LEARNING OF PARTS AND WHOLE OBJECTS IN INFEROTEMPORAL CORTEX

by

Andrew W. Cheng

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

November, 2020

Abstract

Visual object recognition is among the most vital of abilities required to function properly. Our brains perform this impressive computational feat constantly each moment as we make sense of the world about us. A vast history of work has elucidated the emergence of object recognition along the ventral pathway in cortex. The inferotemporal cortex, at the end of the ventral pathway, represents the highest level processing. Single neurons here are able to capture entire objects and faces. Furthermore, IT representations are plastic. Throughout our lives we are able to gain or hone visual discriminatory ability through experience and repetition, and this is driven by the ability of IT neurons to enhance its selectivity of behaviorally relevant objects.

While it is clear that complex stimuli reliably drive IT activity, one relatively unexplored topic is whether that activity indicates holistic representations of the exact stimulus presented, or alternatively, a combinatorial representation of multiple stimulus subparts. The current view in the field is the former. Past evidence, to be reviewed here, suggests that learning strongly enhances selectivity to part combina-

tions. This effect is stronger than the enhancement of selectivity to individual parts, and furthermore, the part combination enhancement is manifest in a way that boosts the sparseness of neural coding: singular objects evoke singularly high responses. Taken together, the widely accepted view is that IT neurons code objects holistically.

This project readdresses this question with an extended experimental design relative to previous work on the topic. A novel set of control stimuli with a wider range of combinations is employed to address a key limitation in the previous design. By directly dissociating and testing holistic versus part selectivity tuning in a way not possible before, we report counterevidence of holistic representation. This demonstrates that our previous reports of compositional coding in IT persists even through learned experience.

We also further characterize the learning effect by exploring the shape tuning of IT neurons, some with and some without a learning effect, with a medial axis mathematical model. We show that shape representation in unlearned cells are biased to top sections of stimuli, while learned cells equally represent top and bottom sections.

Primary Reader: Charles E. Connor

Secondary Reader: Kathleen E. Cullen

Acknowledgments

I first entered graduate school with a fundamentally individualistic mindset. All that I wanted to accomplish, I wanted to accomplish it alone through sheer talent and willpower. I leave grad school with a very different outlook. Much more than the work, I deeply value the people I have met along the way, as well as the people who have been there all along. This accomplishment truly belongs to the love and support around me, and this is the most profound discovery I have made during my wonderful years at Johns Hopkins and Baltimore. Thank you to my beautiful Kelly. You have patiently supported me and believed in me all these years. There is so much to say, I can't think of a single idea to encapsulate it. Just know that I will convey my gratitude to you with my actions. To Mom, Dad, and Jo: well you guys have been waiting even longer! I've never known anything but unconditional love from you all. I don't take that lightly. I promise to call back more! To Kelly's wonderful family: your hospitality and care over the years amazes me. I can't imagine what life would have been like if you hadn't treated me like your own son. I am forever grateful. To Ed, it has been a pleasure working for and learning from you. I have

always admired the lab environment you foster and the kindness and commitment you display to your students and coworkers. To Steve: I cherish the short amount of time that I knew you. I am still sad I never got to finish our project, but hopefully I will stlll get the chance to honor our legacy someday. To the amazing support system at MBI over the years: Justin, Ofelia, the Bills, Hao-Lei, Soohoo, Charlie, Kristiana, Austin, etc...you guys are truly the unsung heroes. You spoil us grad students like no other! And to all my close friends and labmates over the years: this experience has truly been special. Because of you all, I lived in a cozy bubble all these years for better or worse. You made Baltimore the "Greatest City in America".

Dedication

To all my Family. This is for you.

Contents

List of Tables

List of Figures

Chapter 1

Background and Motivation

1.1 Visual Object Recognition

One of the most crucial and complex abilities we possess is visual object recognition. We live in a fast-paced, dynamic environment, but we are able to detect and recognize meaningful objects within them and act upon this information. This has afforded us such abilities as identifying predators and prey, to recognizing edible versus poisonous foods, to diagnosing tumors in medical images, to developing written language and symbols, to finding familiar faces in a crowd. These abilities come so naturally and effortlessly to humans and other visual primates that it can be easy to take them for granted. But the challenges of visual recognition are computationally vast and an understanding of how our brains manage this task would have far reaching implications [1].

1.1.1 Ventral Pathway

Neural processing of vision is thought to occur along parallel streams. The ventral and dorsal pathways are nicknamed the "what" and "where" pathways respectively [2]. The dorsal stream is thought to be involved in recognition of where objects are in space as well as action guidance. The ventral stream, of which this project is concerned, is involved in object recognition [3]. A wide range of methods have shown that neuronal processes supporting object recognition are located in the ventral pathway. Ventral pathway has a rough hierarchy of cortical processing stages. In monkeys these stages are V1 (primary visual cortex), V2, V4, and inferior temporal cortex (IT), which itself is separated into posterior (PIT) and anterior (AIT) cortices. In humans, the lateral occipital complex correspondent to AIT in monkeys [4].

Information flowing up the ventral pathway gets transformed into successively more complex abstractions [5, 6]. At early processing stages, objects are represented by their constituent parts. Neurons here code for simple properties of local edge fragments like position, orientation, and curvature [7–10]. With simple and local information represented at each neuron, objects are thus encoded by large, highly distributed patterns of neural activity [11]. This is in contrast with the late stages of ventral pathway processing.

1.1.2 Inferotemporal Cortex

IT cortex, sitting at the end of the ventral pathway, exhibits the most complex object representation. Single neurons in IT cortex have receptive fields covering major portions of our field of view, typically overlapping the fovea and crossing hemifields [12–14]. These large receptive fields are able to capture entire objects and faces [15]. Furthermore, the receptive fields are more complex than the lower areas. Stimuli that reliably drive cells in primary visual areas are less effective in IT [12, 16–18], as they are generally too simple and pertain to specific foundational aspects of the visual field. Instead, IT neurons reliably respond to complex shape features or whole objects and faces [15, 19–23], while exhibiting complex features such as size and position invariance [14, 24, 25].

The shift from representation of constituent parts in earlier processing areas to entire objects in IT also has implications for how information is stored on a population level. Unlike the distributed patterns of neural activity of previous areas, IT neurons encode objects in a sparser pattern [12,23,26–30]. This sparser pattern is furthermore not organized as previous areas are. IT is the first area that does not display a clear retinopy [6].

Beyond demonstrating the presence of object recognition processing in IT, other work has demonstrated that IT activity is necessary for object recognition. Lesions or disconnections in IT have demonstrated the importance of IT for object [31–33] and face [34, 35] discrimination. Conversely, stimulation of IT neurons can influence

perception as has been demonstrated in face-selective areas [36].

1.2 Role of Learning in Object Recognition

A major component to object recognition is the role of learning. In the wild, vervet monkeys naturally "categorize" potential predators, producing unique vocalizations in response [37]. In the lab, the role of learning has been studied through developed visual systems [38], visual systems that have been deprived of early-life experience [39,40], and the role of visual experience in adults [41,42]. Studies in the pigeon were the first to demonstrate generalization of categories of images such as "human" and "non-human" [43,44].

For the purposes of this project, we focus on experience-based plasticity in the adult visual system. Throughout our lives we are able to gain or hone discriminatory ability through experience and repetition [26,27,45]. In doing so, we train our visual system to categorize and discriminate between behaviorally relevant stimuli in ways that a naive brain cannot do. Many psychophysical studies in adults have shown learning-dependent changes in discriminiation and recognition using stimuli ranging from simple features, such as oriented lines and gratings [46], to complex objects [47]. When looking at neural changes in response to learning, it has been reported (in primary visual areas) that more cells respond to the trained location,

indicating a "swelling" of the cortical map of response preferences. In other areas of cortex, examples of cortical reorganization or magnification have been reported in somatosensory and auditory areas respectively [48, 49].

1.2.1 Learning in the IT cortex

For IT cortex, the change in neural activity in response to learning is most frequently reported as a sparsening of activity. This is often seen as a sharpening of a cell's tuning curve. Furthermore, the sharpened tuning curve tends not to be a result of an increased maximum evoked firing rate, but rather a decrease in sub-optimal evoked firing rate. That is, in general the optimal preferred stimulus pre-training stays constant and the evoked firing rate stays constant while sub-optimal stimuli evoke less response post-training relative to pre-training. This effect has been reported for simple stimulus characteristics such as orientation discimination [50, 51] and random-dot coherent motion [52], as well as complex objects. To illustrate further the observed learning changes with complex objects, some example studies are highlighted.

1.2.2 Highlighted Study: Kobatake 1998

For the first highlighted study, 45, experimenters trained two monkeys to recognize 2-dimensional, monochromatic black images. Monkeys were tasked to remember

individually presented sample objects and, after a delay period, select the sample object when it reappeared on the screen along with distractors. Both the sample and distractors were drawn from a pool of 28 training stimuli. Each stimulus was composed of an arbitrary set of lines, ovaloids and polygons overlaid on top of each other to create one single object. All colors were monochromatic black. Monkeys took roughly 3-5 months of training 8 hours per day, 6 days a week to reach 75% performance.

After training, IT neurons were recorded from the trained monkeys under anesthesia as images were flashed passively. The images were composed of all the training stimuli as well as 75 reference stimuli. The reference stimuli were composed of pictures of random objects found around the lab. All images were in black and white. In addition to the two trained monkeys from which 131 neurons were recorded from, there were three control monkeys who were not trained on the task, out of which 130 neurons were recorded from. The existence of control monkeys adds extra power to this project's results, as it is uncommon for projects to have the resources to record from completely naive monkeys.

The experimenters first considered the best elicited responses (called "maxFr") across cells and they found a modest training effect. Responses across trained stimuli show a learning effect of increased maxFr. The distributions of maxFr from the 131 trained cells are significantly different from the distribution of the 130 untrained cells (Kolmogorov-Smirnov test, p¡0.01). In contrast, when considering responses to

reference stimuli, there is no significant difference between the same distributions.

Next, they considered the sparseness of coding as ascertained by stimulus discriminability. They estimated the population coding of stimuli by constructing response vectors of the training and control cells. A training response vector for a particular stimulus is a 131-length vector with each entry being a normalized training cell response for that stimulus. Control response vectors are constructed the same way, with length 130. From the set of training and control vectors, stimulus distance is defined as the vector difference between any two stimulus response vectors (of the same type: training-vs-training or control-vs-control). Finally, the sparseness of coding is then indirectly ascertained by considering the distribution of all possible stimulus distances across training and control sets. When doing so, they found a significant training effect. Stimulus distance across training cells were significantly greater than stimulus distance across control cells (KS test, p¡0.001) and importantly, this effect was far greater than the maxFr training effect. This supports the view that training sharpens a cell's tuning curve (thereby increasing stimulus distance and discriminability), while keeping the optimal evoked response constant (they actually report a modest increase, but the increase does not match the increase in stimulus distance).

1.2.3 Highlighted Study: Freedman 2003, 2006

The next results we consider are from a pair of studies: Freedman 2003 and Freedman 2006 [27, 53]. In both studies, experimenters created a shape space consisting

of virtual models of cats and dogs. Each stimulus was a clearly-defined virtual 3D representation of a cat or dog. They defined three canonical models of cats and dogs each (adapted from three actual species). They then defined a morphing space between canonical models. They could take any two canonical models – cat1 and dog1, for example – and parametrically define any image within the continuum space between 100% cat1 and 100% dog1. Each mixture image was still a sharp, well-defined shape. Images consisting of over 50% of any canonical model was considered part of the category of that canonical model. Furthermore, images could be categorized into a broader cat category versus dog category (by having more than or less than 50% of any cat model respectively).

Monkeys were tasked to categorize cats versus dogs within a delayed match-to-category task. The sample image, shown first, was drawn from one of the six canonical images. The test image, shown second and after a delay period, was a morphed image generated from anywhere in this shape space. If the two displayed categories matched, monkeys responded by pressing a lever. In the case of a nonmatch, monkeys had to wait through another delay period and get to a second test period, in which another morphed stimulus was presented. This second test stimulus was always a match with the sample. This atypical experimental design is due to the experimenters recording from both IT and prefrontal cortex. For the purposes of this review, the main result of note is that across neurons, evoked responses from the sample, first test epochs of neuronal response displayed categorical representation. Representations for the

three cats tended to be similar as did the representations for the three dogs. Intra-category variation was smaller than inter-category variation. (Prefrontal cortex was shown to have greater categorical information than IT and so the paper argues that categorical information is mainly a feature of PFC over IT. Nevertheless, they do show categorical information present in IT and once again, the PFC considerations are beyond the scope of our purposes).

In a follow up study [53], experimenters modified the experiment to investigate learning effects in IT. They considered 18 stimuli from the previous experiment. The 18 stimuli consisted of 3 different morph lines (cat1 to dog1, cat2 to dog2, and cat3 to dog3) each with 6 levels of morph (from 100% cat to 100% dog). Of these stimuli they generated copies at seven different orientations. 0 degrees rotation correspond to the set of original stimuli, and these were considered "trained" stimuli. The stimuli at the other six orientations (at 22.5, 45, 67.5, 90, 135, and 180 degrees) were considered novel (untrained).

The results of this study showed clear and striking evidence of the aforementioned learning effects in IT. They found the sharpest tuning amongst the trained stimuli. That is, tuning was sharp among the 18 trained stimuli (0 degree orientation) and furthermore, the sharpness of tuning among 18 stimuli of any other (non-trained) orientation was not as high as that of the trained stimuli. In fact, with increasing orientation there was a clear monotonic decrease in tuning sharpness. However when comparing the maximum evoked response of each orientation with one another, there

was minimal difference across the orientations. So once again, experimenters found that training increased the sharpness of tuning without increasing the maximum evoked response of the optimal stimulus.

1.2.4 Limitations

In these highlighted studies, along with most other previous work in the matter, it is clear that a learning effect is taking place. However, because the stimuli that drive IT neurons are so complex, it is not immediately clear what the neurons are learning. Each of the stimuli in the Kobatake and Freedman studies have multiple parts (for example: crosses and ovaloids for Kobatake, heads and tails for Freedman). From the design of these studies, it is impossible to know whether the sharpened tuning resulting from training is manifest through modulated sensitivity to a specific sub-part of the optimal stimulus (i.e. the shape of the tail of a cat), or multiple parts of the optimal stimulus, or the entire optimal stimulus itself. In fact, in the vast body of work concerning learning in IT, only one major study addresses this question.

1.3 Baker, Behrmann, and Olson 2002: Learning Parts Versus Wholes

In 2002, Chris Baker, Marlene Behrmann, and Carl Olson [26] released a seminal and widely impactful paper that directly investigates whether IT neurons code for object parts or whole objects. Their main conclusion of whole-object coding has been widely accepted and has deep implications for object representation in cortex. The work reported in this thesis is a direct successor of their study.

1.3.1 Experimental Design

Experimenters constructed their stimuli, called batons, by joining two distinct top and bottom elements by a vertical stem (Figure 1.1a). There were 16 total batons constructed organized into 4 groupings called tetrads. Each tetrad consisted of 4 batons, which shared amongst them two top parts and two bottom parts combinatorially organized into a 2x2 grid. Monkeys were trained to distinguish batons by depressing a left or right lever (denoted in Figure 1.1a, grey and white backgrounds respectively). The batons sharing the same response type (right or left lever) had no parts in common.

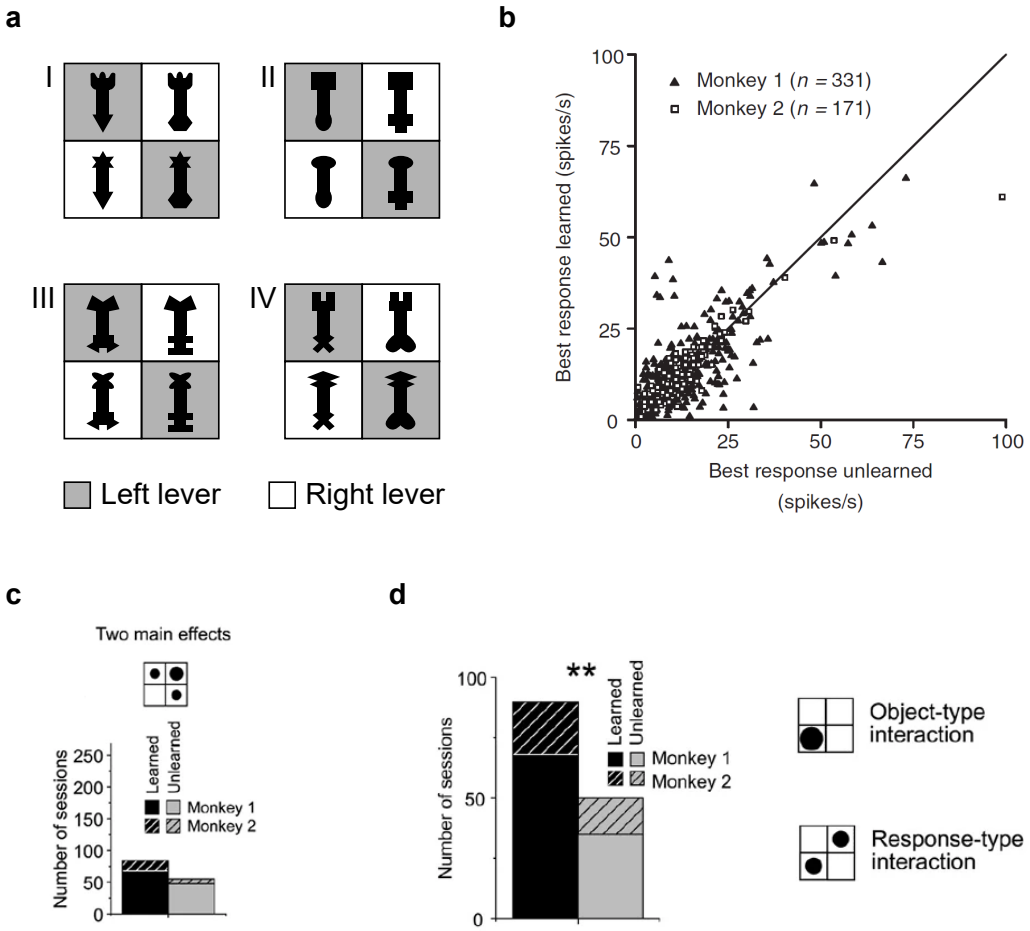
Two monkeys were trained on the task, with each monkey trained on two tetrads. The trained tetrads were different and complementary among the monkeys (Monkey 1 was trained on tetrads I and II, Monkey 2 was trained on tetrads III and IV).

Figure 1.1: Baker et. al. 2002: Batons and Results

Stimuli and results from Baker et. al. 2002 [26]. a) Four tetrads of batons were used in discrimination training. Monkey 1 was trained on tetrads I and II and monkey 2 on tetrads III and IV. The batons used in training for one monkey were also used as unlearned controls for the other monkey. Batons requiring right- and left-lever responses are indicated by white and gray backgrounds, respectively. However, during experiments, the background was constant. b) Response to the best learned baton plotted against response to the best unlearned baton. Each point represents data from one session, where each session includes responses of one neuron to four batons of a learned tetrad and four batons of an unlearned tetrad. Each neuron provides up to two sessions. There was no significant tendency in either monkey for responses elicited by the best learned baton to exceed those elicited by the best unlearned baton (paired t-test: monkey I, $p > 0.7$; monkey 2, $p > 0.5$), c) Selectivity for the individual parts of learned batons was enhanced relative to selectivity for individual parts of unlearned batons. A two-way ANOVA with top part and bottom part as factors was executed. The four histogram bars represent counts of sessions in which neurons showed two significant main effects of part identity for batons belonging to learned (black) or unlearned (gray) tetrads in monkey I (uniform texture) or monkey 2 (hatched). The sessions displaying at least one learned main effect was significantly greater than the number displaying at least one unlearned main effect ($p < .0005$, Chi-squared test) d) Learning enhanced the tendency of neurons to respond selectivity to specific part combinations. The histogram bars show the counts of sessions in which the two-way ANOVA yielded evidence of a significant nonlinear interaction between the influences of top and bottom parts. For learned tetrads (black) as compared to unlearned tetrads (gray), significantly more interaction effects occurred ($p < 0.0001$, Chi-squared test).

(next page)

Figure 1.1: Baker et. al. 2002: Batons and Results



1.3.2 Recording

For data collection, awake monkeys underwent maintain-fixation tasks. Neurons were first screened for visual responsiveness, and then significant responsiveness to at least one of the 16 batons. After screening, a subset or all (time permitting) the batons were redisplayed for 8 trials each, and this constituted the experimental data that was analyzed. The subset of batons that were prioritized were four batons from the one trained tetrad and four batons from one untrained tetrad. Both of these tetrads were selected for having a higher maximum elicited response than the highest elicited response of the other corresponding tetrad. That is, the selected trained tetrad had a higher maximum elicited response than the maximum elicited response of the other trained tetrad, and the same for the selected untrained tetrad. So in total, eight batons were prioritized over the other eight. The data from these batons (spanning trained and untrained data) constituted one "session". If time permitted, experimenters then displayed the other eight batons (which were from the other trained and untrained tetrad), and the data collected constituted another session. Each recorded cell provided either 1 or 2 sessions of data and each session was treated as an independent observation. They recorded from a total of 360 cells out of which, 142 had two sessions and the rest had one session for a total of 502 sessions.

1.3.3 Results and Interpretation

Similar to previous studies, they found no learned effect that boosted the maximum elicited response of trained tetrads over untrained tetrads (Figure 1.1b). When comparing the best responses from trained versus untrained tetrads (from the same session), no significant difference was found (paired ttest; monkey 1, $p > 0.7$; monkey 2, $p > 0.5$). Once again, this is consistent with the existing understanding of learning in IT.

To assess the learned effect on selectivity, they carried out two separate 2-way ANOVAs. for each session. One ANOVA was applied to the trained tetrad while one was applied to the untrained tetrad. 2-way ANOVA yields main effects and interaction effects. The "main effects" in this case directly report "part selectivity". That is, when applying 2-way ANOVA across responses from any particular tetrad, the resulting F-values of "top" and "bottom" main effects are a direct measure of the discriminability of the two top and bottom parts of that tetrad. The "interaction effect" of 2-way ANOVA, however, is more complicated, as multiple factors could contribute to an interactive effect. More on this will be discussed later. For assessing both main and interaction effects across the population, the experimenters first assessed the significance of both main and interactive effects for each tetrad of each session (significance was ascertained by the standard 0.05 p-value threshold. The p-values are from the ANOVA calculation of the F-values). The significance results are then pooled across sessions.

1.3.3.1 Part Selectivity

Figure 1.1c and 1.1d show the ANOVA results of the pooled main and interactive effects respectively. For the main effects, each histogram shows whether the ANOVA results had zero, one, or two significant main effects. Results from trained and untrained tetrads are differentiated and compared, as well as results from individual monkeys. Pooling results from both monkeys, the experimenters reported a higher proportion of responses from trained tetrads exhibiting one or two main effects as compared with untrained tetrads. Using a Chi-squared test, they find this effect to be significant ($p < 0.0005$).

1.3.3.2 Part Interaction Selectivity

For part interaction, they found a similar but more pronounced effect. Once again, a higher proportion of trained tetrads had significant interaction effects than untrained tetrads. Furthermore this proportion disparity was even higher than the disparity reported for main effects. Trained tetrads provided almost twice as many significant responses than untrained tetrads, and this was reflected in a stronger chi-squared test ($p < 0.0001$). This result is striking, but before further interpretation, the interaction effect itself must be parsed.

1.3.4 Whole Object Selectivity

All interaction effects, by definition, are deviations from the linearly summed main effects. There are multiple ways in which these deviations can manifest. Firstly, response from a single baton could exhibit large deviation with respect to the rest of the tetrad. This type of interaction was called object-type interaction (Figure 2.1a). Secondly, two diagonal batons sharing no parts in common could have the large (same-direction) deviations (Figure 2.1b). This would lead to an XOR-gate scenario unexplainable by linearly summing two main effects. They label this "response-type" interaction because of the nature of the experimental design: it would indicate that a neuron tuned to the response type (left or right lever) that the monkey had to deliver as opposed to being stimulus-selective. To differentiate between these two types of interaction effects, they devised an index based on the range and variance of the response across the tetrad. Relevant edge cases include an index of zero for a perfect response-type interaction, four for a perfect object-type interaction, and six for a no interaction effect (where responses are perfectly described by linear summing main effects).

Using this index, they could cast the response of each tetrad onto a continuous range from zero to six (Figure 2.1c). What the graph shows is that 1) Interaction effects are indeed prevalent across most tetrads (the vast majority of tetrads exhibit an index between zero and four, indicating some part interaction), and 2) learned tetrad part interaction, relative to unlearned part interaction, is shifted toward object-type

interaction and away from response-type interaction. Thus, with learning comes not only increased part interaction as a central effect, but specifically a type of part interaction that boosts selectivity of a single baton (among the tetrad). The authors interpret this as evidence that IT neurons learn to select for entire objects, and not just for object parts or combination of parts. This is the key interpretation of this study, and it has been widely accepted and enormously influential in the field. It implies that, say in the aforementioned Freedman studies, that some IT neurons are selective for a specific type of cat – not a combination of its tail and head, but rather the entire cat – and learning functions to improve this selectivity and cancel the noise. Indeed, this interpretation seems well supported by the evidence provided and furthermore the interpretation fits neatly with the prevailing notions of learning in IT. Object-type interaction as they define it is very closely related to sparseness (a response profile that is perfectly sparse would also exhibit perfect object-type interaction), and so the result that learning boosts object-type interaction fits perfectly with the notion that learning increases the sparseness of coding in IT.

1.3.5 Recap

To recap, Baker, Behrmann, and Olsen implemented a novel design to be able to directly manipulate stimulus parts in a combinatorial fashion. They found a learning effect on part selectivity, and an even greater learning effect on part interaction selectivity. Furthermore, they show that the increased part interaction selectivity is

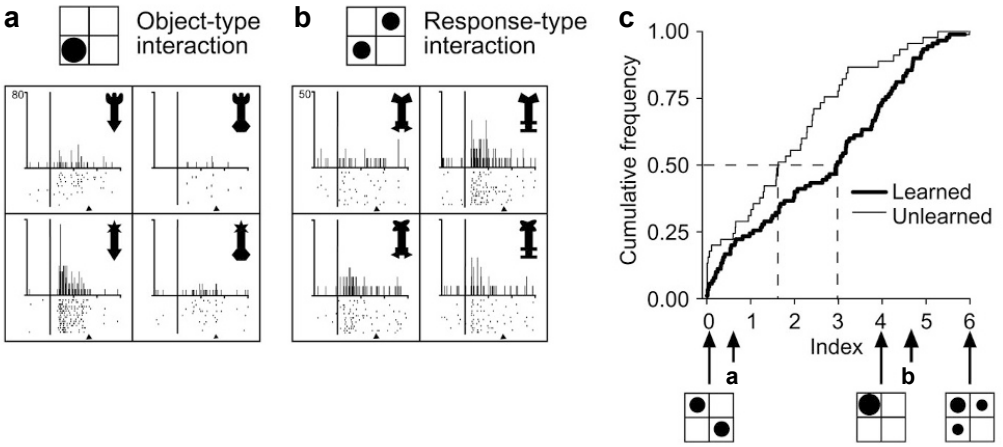


Figure 1.2: From Baker et. al. 2002 [26]. Learning is characterized not only by the enhancement of part interaction selectivity, but specifically the enhancement of supra-additive interaction effects of single objects within the tetrad. (a,b) Neurons showing significant interaction effects occupied a continuum extending from 'object-type' cases (example in a) to 'response-type' cases (b, batons sharing no parts, but associated with the same behavioral response, elicited equal responses). (c) Cumulative frequency of an index developed to parse between different types of interaction. Index = $(x_1^2 + x_4^2)$ / V , where x_1 and x_4 are the firing rates elicited by the best baton and the baton sharing no parts with it respectively and V is the variance across the evoked responses of the four batons. The curve for learned tetrads (thick) is shifted relative to the curve for unlearned tetrads (thin) away from 0.0 (the value associated with a pure response-type pattern) and toward 4.0 (the value associated with a pure object-type pattern). The value of 6.0 is associated with no interaction at all. The shift of the learned curve is significant ($p < 0.01$, Kolmogorov-Smirnov test). Index values for neurons in (a) and (b) are indicated by arrows.

specifically object-type (single baton) selectivity, and they use this as evidence to conclude that IT neurons code for whole objects as opposed to object parts.

1.3.6 Problems

However, there is a major drawback with the experimental design of Baker et. al. While they show that the learned increase in part interaction selectivity manifests in increased sparseness coding, which in turn can be seen as compelling evidence for whole object selectivity, the conceptual link between part interaction selectivity and whole object selectivity is nevertheless intuited and not firmly established. Their experimental design does not allow them to directly test this link further in that they cannot distinguish between a general increase in part interaction selectivity and whole object selectivity. The work presented in this thesis will address this issue.

1.4 Objectives

The work described in this thesis is a direct successor to the results of Baker, Behrmann, and Olsen. The goal of this research is to extend their findings in several key ways. We separate the work into two specific aims.

Specific Aim 1: Test the hypothesis that learned visual recognition enhances whole object selectivity in IT. We describe a new experimental design which incorporates new unlearned controls in order to address the limitations de-

scribed above. We also detail other improvements that we believe provide a stronger and more-complete view of the neural changes taking place. As a result of these improvements, we first report a confirmation of the main findings of Baker et. al., but then we are able to extend the results and provide counterevidence to whole object selectivity. We conclude that parts-based compositional coding, which we have previously reported, remains the fundamental feature of IT activity, even throughout learning.

Specific Aim 2: Analyze the shape tuning changes that explain whole object selectivity. Here, we endeavor to precisely describe neuron tuning following learning. Using a genetic algorithm, we explore the shape space of IT neurons and constrain a complex mathematical model. Using this model we compare differences between cells that exhibit learning effects versus cells that do not. Our efforts provide a modest result that information carried in unlearned cells are biased towards (spatially) top stimulus parts, while learned cells exhibit equal information representation across familiar top and bottom stimulus parts.

Chapter 2

Parts-Based Compositional Coding In Learned Object Recognition

2.1 Motivation

The results and interpretation of Baker et. al. have been widely influential in the field of visual object recognition. Cited by over 300 articles, their findings have been widely accepted to date. The assertion of whole-object coding, furthermore, has strong implications for cortical object representation. It implies that each object we learn throughout our lives is ultimately represented by a set of dedicated neurons. These dedicated neurons would not be responding to parts or multi-part combinations, but the entire object. From an information standpoint, this would necessitate investigation into how IT cortex could handle the task of representing every single

behaviorally relevant object through labeled line coding.

However, their work leaves open another possibility. The familiarity of parts may be sufficient to produce the increased selectivity they report. While it is striking that they report that learning specifically enhances supra-additive selectivity for singular objects, the link from this finding to the conclusion of holistic coding is nevertheless inferred. A neural code based on part selectivity (which can include nonlinear combinations of constituent part response) could also produce the observed supra-additive profiles. Their experimental design was novel and powerful, but as discussed in the previous chapter, it was not equipped to determine which coding scheme (parts or holistic) produced the supra-additive profile,

In this chapter, we introduce an improved experimental design that is able to directly dissociate between parts and holistic coding. We detail this along with other improvements.

2.2 Methods

2.2.1 Training Stimuli

We defined 32 letter-like two-dimensional stimulus categories. Each category was defined by its medial-axis topology. Each stimulus category consisted of a "top" and a "bottom" section. Figure 2.1a displays the first 16 categories. Four tops and four bottoms comprise the 16 categories in a combinatorial fashion. We trained 2 monkeys

on complementary halves of the stimuli. Irrespective of monkey, the set of 8 trained categories is referred to as the "Trained" set. The other 8 categories are referred to as the "Complementary" set. "Complementary" categories consist of tops and bottoms familiar to the monkey (the same tops and bottoms compose "Trained" categories), but in unfamiliar combinations (thus "combination-Complementary" stimuli). Collectively, all 16 categories are referred to as the "Canonical" set. A set of 16 entirely unfamiliar comparison categories (unfamiliar parts and unfamiliar combinations) was constructed by inverting (rotating 180 degrees) the canonical set (Figure 2.1b). This set of categories is referred to as the "Inverted" set.

For each category we could generate morphed versions by varying lengths, widths and smoothness of component limbs (Figure 2.1c, see section 2.2.2 Stimulus Generation for more detail), but limb orientation remained the same. The monkeys were tasked to differentiate and match whole categories, not single images. The original unmorphed images of each category, as shown in Figure 2.1a and 2.1b, are referred to as "Unmorphed" stimuli. All morphed versions of the categories are referred to as "Morphed" stimuli.

The terms "Trained", "Complementary", "Canonical", "Inverted", "Unmorphed", and "Morphed" will be used either in the context of individual stimuli, sets of stimuli, or stimulus responses.

Figure 2.1: Methods: Experimental Design

(a,b) The stimuli categories of our experimental design. Inspired by Baker et. al., we designed 32 letter-like stimuli in a combinatorial manner with shared tops and bottoms. Tops are shared within the same columns and bottoms are shared within the same row. Canonical stimuli are shown in (a). Two monkeys were trained on complementary halves of the Canonical Stimuli, denoted by the green and violet backgrounds. The halves are referred two as Trained and Complementary sets of stimuli. Complementary stimuli are unlearned controls. Shown in (b) are Inverted stimuli, which are flipped versions of the Canonical stimuli. No monkeys were trained to discriminate these Inverted stimuli, and so they are also unlearned controls. Each of the stimuli shown in (a,b) are Unmorphed versions of a category of stimuli. We trained monkeys to distinguish between categories of stimuli while considering any Morphed versions of any category stimuli to be the same. Shown in c) are example morphed versions of a single category. Limb lengths and widths are varied, but limb orientations remain the same. (d,e) show the batons of Baker et. al., but recast into the organization of our stimuli. d) shows learned batons (corresponding to our Canonical stimuli) while e) shows unlearned batons (corresponding to our Inverted stimuli). Note that the significant addition to our design is the Complementary stimuli, which are composed of familiar (trained) letter parts, but novel part combinations. f) A schematic showing the advantage of Complementary stimuli. Learning would differentially affect responses of Complementary stimuli from a whole-object-selectivity context versus a general part interaction selectivity context. With whole object selectivity, learning would only enhance selectivity among Trained stimuli, whereas with part interaction selectivity, learning could enhance both Trained and Complementary stimuli. Without these controls, Baker et. al, could not distinguish between whole-object and part-interaction selectivity.

(next page)

Figure 2.1: Methods: Experimental Design

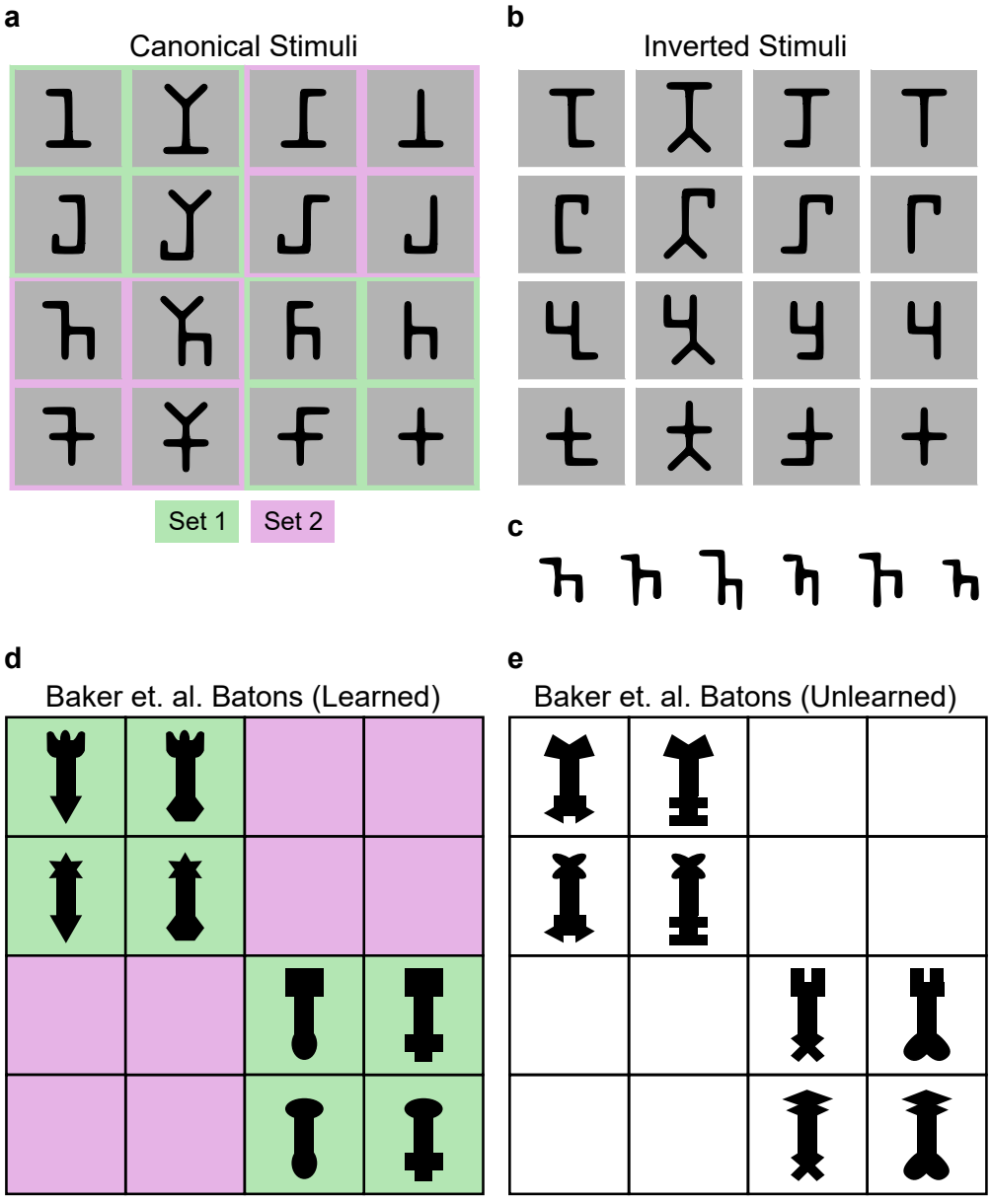
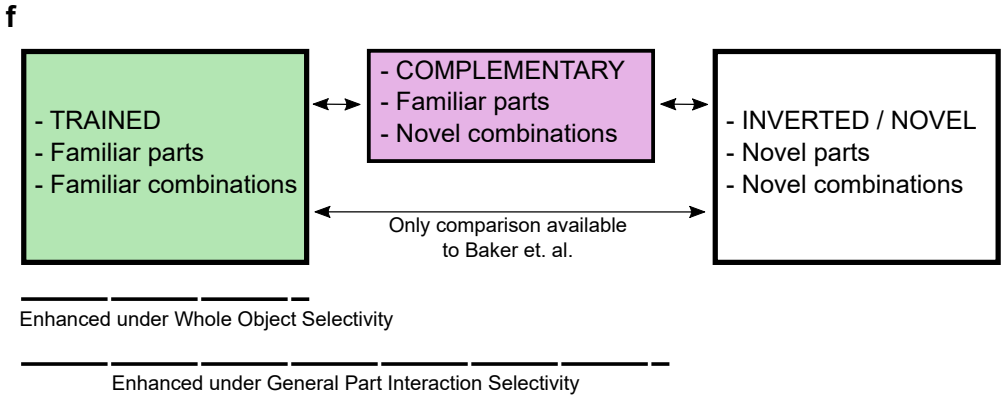


Figure 2.1: Methods: Experimental Design (Cont)



2.2.1.1 Comparison to Baker et. al. 2002

Our stimuli are inspired by Baker et. al., but there are important differences. We tasked our monkeys to recognize and differentiate between stimulus categories instead of single images. In this way, we could ensure that monkeys learned generic shapes, and not rely on ad hoc strategies (involving precise, local details). This is similar to the way we learn alphanumeric characters, which vary in precise shape across fonts and handwriting styles.

In Figure 2.1d and 2.1e, we recast the batons from Baker et. al. into the same organization of our stimulus categories. Same as Baker et. al., our Trained set could be organized into directly comparable "tetrads" in that there are two groups of four trained categories sharing two tops and two bottoms (however the tetrad organization is not used in our analysis). Our Inverted set is directly comparable to the untrained batons of Baker et. al., in that they were composed of unfamiliar parts and unfamiliar

part combinations. However, our Complementary set, composed of familiar parts but unfamiliar part combinations is unique to our study. This afforded us the ability to directly test changes in part combination selectivity (in the context of ANOVA analysis, called "part interaction" selectivity) in a way that Baker et. al. could not.

Figure 2.1f conceptually establishes what is gained by the inclusion of the Complementary set: it allows for a dissociation between general part interaction selectivity and whole object selectivity. Whole object selectivity and part interaction selectivity would differentially affect the responses of the Complementary set. In the context of whole object selectivity, a learned increase in spareness of coding must by definition involve familiarity with an entire object. Since the only familiar entire objects occur in the set of familiar stimuli, Complementary stimuli would not be "selected for" following holistic learning. By contrast, in the context of general part combination selectivity, only familiarity of parts is necessary to be influenced by learning. Therefore, selectivity of Complementary stimuli could theoretically be enhanced following learning. Thus, the inclusion of Complementary stimuli is a powerful addition to the experimental design.

2.2.2 Stimulus Generation

For every stimulus, a medial axis skeleton is first generated. "Limb" precursors (where each precursor is a width-less line segment defined by the endpoint locations in two dimensions) are joined together at their endpoints to form a stimulus skeleton.

The skeleton must be fully connected (So all the limbs topologically form one stimulus, not multiple) and no loops are allowed (no "triangle" or "rectangular" shapes which have closed loops. Topologically, everything must be a "tree" with branches and endpoints). The endpoints of all the limbs in the skeleton are considered nodes.

From the skeleton, a surface contour is then built. Each node is assigned a width. The width is roughly a "radius" about the node by which a set of "control points" are generated. Control points are surface contour precursors, roughly tracing the outline of the final shape. For terminator nodes, three or four control points were generated about the node (three-control-point nodes generate "smoother" variants while four-control-point nodes generate a sharper convexity). The width assigned to the terminator determined the average distance between control points and node. For junction nodes, two or three control points were generated for each concavity region between the limbs sharing the node. For example, a node joining three limbs would have 6 or 9 control points: 2 or 3 control points for each of the three concavities between limbs. Finally, the set of all control points are used to generate a spline curve to form the final closed surface contour defining the shape.

For the Behavioral stimuli (Unmorphed stimuli categories), all the node widths are uniform as well as the smoothness (four nodes about terminators, 3 nodes about junction concavities), leading to a smooth, regular shape. Also the limb lengths and junction angles are defined and at regular repeated angles.

Morphed stimuli are morphed versions of the Behavioral stimuli. The generated

stimuli start from the same skeletons of the Behavioral stimuli. From there, certain parameters are changed (morphed). All limb lengths, node widths, and smoothness at each node were changeable parameters, and changed at random. Importantly, junction angles were kept constant so that the general shape characteristics kept constant (no Morphed stimuli could be mathematically confused for a different category). The continuous parameters of limb lengths and node widths were assigned a value range that they could morph into. All Morphed stimuli were checked for consistency (making sure that a final smooth contour was possible), before being finalized.

2.2.3 Protocols

2.2.3.1 Training Protocols

The training paradigm was a standard match to category task (Figure 2.2a). Following fixation on a center dot, a Morphed Trained stimulus was presented for 500 ms. A second, Unmorphed Trained stimulus was presented after a 500 ms delay. If the two stimuli were a categorical match, the monkey could earn a liquid reward by making an upward saccade. If the two categories did not match, the monkey was rewarded for maintaining fixation at the center. The task was difficult because of the confusability of morphing, letter-like stimuli constructed from the same set of component parts. As a result, performance on any given stimulus category only gradually increased from chance to the 80% criterion, over the course of 20 to 40 training days (Figure 2.2b).

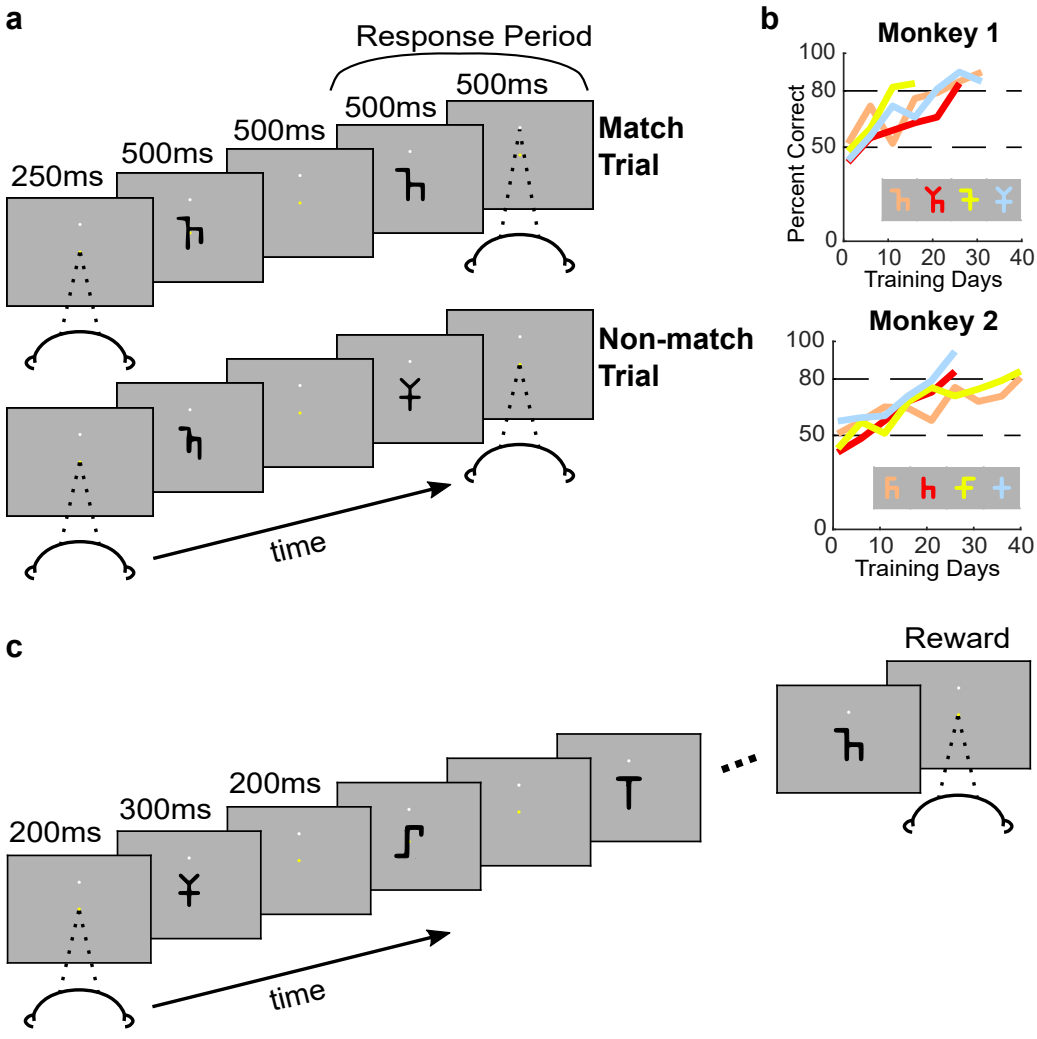
New stimulus categories were introduced at intervals of several weeks, but there was no substantial change in learning rate across successively introduced categories (the curves in Figure 2.2b are aligned at the beginning of training for each individual category). This demonstrates that task performance depended on extended learning of each individual shape.

Figure 2.2: Protocols and Training Performance

a) Two monkeys were trained on a standard match to category task. Following fixation on a center dot, two Trained stimuli were presented in sequential fashion for 500ms each with a 500ms delay between presentations. The first stimuli presented was morphed while the second was unmorphed. A reward was delivered following a saccade up to a target if the two stimuli were a categorical match, or a maintained fixation for a categorical nonmatch. b) The average performance of both monkeys throughout a subset of training. Monkeys were trained in stages, with new stimulus categories to compare amongst introduced sequentially. After performance gradually increased from chance to above 80% (over a course of 20 to 40 training days), a new category was introduced to the task. The curves shown here are overlaid to the beginning of training for each individual category. This task was difficult for the monkeys to master, demonstrating the level of expertise needed to execute it successfully. c) The passive maintain-fixation task. Following fixation, eight stimuli were presented in succession for 300ms each with a 200ms interstimulus interval interleaved in between. Fixation must be maintained throughout for a reward. Otherwise the data is discarded, and the trial restarted. Almost all the data presented in this work is recorded during this passive task. The one exception is the Active Task Protocol which a variant of the training protocol in (a).

(next page)

Figure 2.2: Protocols and Training Performance



2.2.3.2 Recording Protocols: Fixation Tasks

Most of the data presented is recorded during a passive viewing task (Figure 2.2c). Following fixation, eight stimuli were presented in succession for 300ms each with a 200ms interstimulus interval interleaved in between. The monkey had to maintain fixation throughout the entire period for a reward. Otherwise the data was discarded, and the trial was restarted. This paradigm is used for a variety of protocols. For all protocols, the set of stimuli shown were randomly interleaved.

2.2.3.3 Behavioral Protocol

This was both the screening protocol as well as a main data collection. All Unmorphed Canonical and Inverted stimuli were shown for at least 15 presentations per stimulus. Cells were only selected for further recording if visual responsiveness (at least two times baseline firing and at a maximum elicited response at least twice as high as the minimum elicited response) was determined. The main results of Aim 1 come from this protocol. Data from this protocol is referred to as "Behavioral responses".

2.2.3.4 Morphed Protocol

For some cells, we presented Morphed Canonical and Morphed Inverted stimuli. A set of five morphs for all 32 categories was generated (only one time). This set of 160 stimuli were saved and the same set was shown to all subsequent cells, five pre-

sentations per stimuli. Data from this protocol is referred to as "Morphed responses"

2.2.3.5 Active Task Protocol

For two of the three recorded hemispheres, we recorded cells while the monkey performed a version of the training task. The stimuli used were still restricted to the set of eight Trained categories. Unlike the training protocol, in which presented categories in each trial are random and the number of trials is indefinite, the recording version had a fixed set of trials and presentations of each stimulus. There were 120 total trials. The first stimulus presentation of each of the trials were determined by drawing randomly (without replacement) from a pool of 120 morphed stimuli. This pool was constructed by taking all eight training categories, generating three morphs for each category, and having five repeats for each resulting stimulus. The second stimulus of each trial is similarly determined by drawing randomly from a pool of 120 unmorphed stimuli, constructed by 15 repeats of the eight categories. The 120 morphed and 120 unmorphed stimuli were randomly matched to generate 120 full trials. No restrictions on occurrences of category pairs or occurrences in match versus nonmatch trials were implemented. During recording, any premature trial break (i.e. the monkey broke fixation before the end of the first stimulus) resulted in an immediate repeat of that trial.

For a small subset of cells (six) in the final recording block (first monkey, second time), the three Morphed stimuli for each category used in this protocol were the

same as the first three (of five) Morphed stimuli used in the Morphed Protocol. That is, the Morphed stimuli were the same across the cells and the same across protocols within the same cell. Comparisons can then be made for responses across cells, and across passive versus active contexts.

This protocol is referred to as "Task". The protocol sharing morph stimuli from the Morph protocol, is referred to as "Task-Matched". Responses from Unmorphed and Morphed stimuli from this protocol are "Task-Unmorph responses" and "Task-Morph responses" respectively.

2.2.4 Electrophysiology Methods

We recorded spiking activity of well-isolated single neurons from two awake monkeys (Macaca mulatta) performing the protocols described above. They were singly housed during training and experiments. All procedures were approved by the Johns Hopkins Animal Care and Use Committee and conformed to US National Institutes of Health and US Department of Agriculture guidelines. Both monkeys were head-restrained and first trained to maintain fixation within a 0.5 degree radius surrounding a 0.25 degree square (fixation spot) displayed on a projection screen 60cm away for a juice reward. Eye position was monitored using a dual-camera,infrared eye tracker (IScan Inc, Woburn, MA).

The electrical activity of 128 neurons (87 and 41 respectively from two monkeys) were recorded with epoxy-coated tungsten electrodes (FHC Microsystems) inserted

through a transdural guide tube. A custom-made chamber was manufactured to provide high precision of cortical targets (within 0.5 mm). Each day, the craniotomy was opened and cleaned, the chamber placed and fastened atop the open chamber with settings adjusted to target a cortical location. A transdural guide tube was lowered 10 to 15 mm into cortex and an epoxy-coated tungsten electrode (FHC) was inserted through the guide tube. Extracellular action potentials were isolated and processed using the TDT RX5 Amplifier (Tucker-Davis Technology, Alachua, FL). Inferotemporal cortex was identified on the bases of the sequence of sulci as the electrode was lowered and visual response characteristics of the neurons. We targeted neurons from the lower bank and lip of the superior temporal sulcus from stereotaxic AP +10 to +21mm, ML 20 to 25 (left hemisphere in Monkey 1, right hemispheres in both monkeys), and DV 0 to +8 (Figure 2.3).

2.2.5 Recorded Cell Breakdown

Protocols were modified during the course of the experiment. Some changes were due to initial oversight, and some were due to natural adaptation as data was analyzed and interpreted. As a result not all cells were shown all the protocols ultimately used in the analysis. Table 2.1 details the breakdown of cells across monkeys and protocols shown. 57 cells across both monkeys were shown both Canonical and Inverted sets in the Behavioral Protocol. These cells are referred to as "Full-Protocol" Cells. Inverted stimuli, and thus Full-Protocol recordings, were only introduced in the sec-

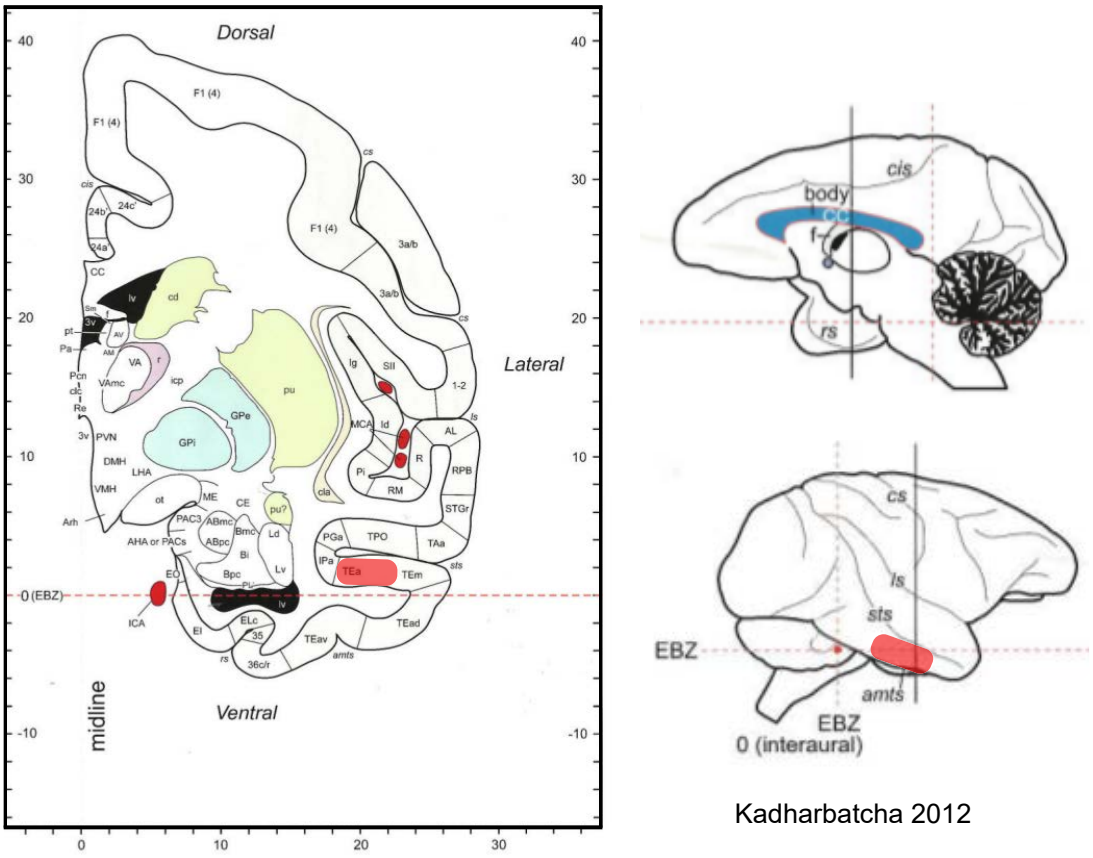


Figure 2.3: Approximate recording locations. Picture is taken from Kadharbatcha 2012 [54] highlighted in translucent red. 128 neurons were recorded from the lower bank and lip of the superior temporal sulcus from stereotaxic AP +10 to +21mm, ML 20 to 25, DV 0 to +8. Both hemispheres of Monkey 1 and the right hemisphere of Monkey 2 were recorded

ond monkey, so they are absent from the first set of recorded cells from the first monkey. After the conclusion of recording in the second monkey, we revisited cells in the first monkey, this time in the right hemisphere to record additional Full-Protocol cells. Full-Protocol cells are used for the analysis in Aim 1, part 1 which compares Canonical responses to Inverted responses.

Within the population of cells in the first recording block (first monkey, left hemisphere), the first 51 cells were only shown Training stimuli in the Behavioral Protocol. Not showing the Complementary stimuli was due to oversight. These 51 cells are referred to as "Train-Only" cells. Their genetic algorithm data is used for Aim 2 of this project. For Aim 1, however, their data is not used (and they are not denoted in Table 2.1).

The remaining 16 cells recorded in the left hemisphere of Monkey 1 were shown Complementary stimuli (but not Inverted). These cells, along with the 57 Full-Protocol cells are labeled "Canonical-Protocol" cells in that they were shown the full Canonical set. These cells are used for the analysis in Aim1-part 2 which compares Trained responses to Complementary responses.

Table 2.1: A breakdown of the number of recorded cells sorted by presented protocols.

	Both Monkeys	**Monkey 1**	**Monkey 2**
All Recorded Cells	128	87	41
Presented: Canonical & Inverted (Aim1, Part 1)	57	20	37
Presented: Canonical (Aim 1, Part 2)	77	36	41

2.2.6 Analysis: Assessing Selectivity

2.2.6.1 Two-Way ANOVA

Selectivity for parts and part combinations was determined by applying two-way ANOVA to the Behavioral responses. Two-way ANOVA is a natural test for our two-factor cross design. For each recorded Full Cell, two separate two-way ANOVAs are applied, one to Canonical responses and one to Inverted responses. This is in the same fashion as Baker et. al. As noted above, the Inverted stimuli are composed of unfamiliar parts and part combinations with relation to the Canonical stimuli, which is the same relationship as the "learned" vs "unlearned sets of baton tetrads in Baker et. al. As such, the results from two-way ANOVAs are interpreted in a similar way. Differences in selectivity to Canonical versus Inverted responses is assessed by comparing the results of the ANOVAs across all Full Cells.

However, it is worth noting some differences between our use of two-way ANOVA and Baker et. al. First, all the Behavioral response data from one cell are treated together. We do not break our responses into separate "sessions" as Baker et. al. did. This gives us a more holistic picture of what each cell is selecting for instead of narrowly focusing on single-tetrad responses at a time.

By applying two-way ANOVA to the Canonical set, we are including the 8 Trained responses, but also the 8 Complementary responses. This could potentially "pollute" the data with unlearned responses. However, because our results later fail to show any

significant differences between Trained and Complementary responses, we conclude that it is fair to consider all Canonical stimuli as the "learned" set for our purposes here.

By including more stimuli, our two-way ANOVA had 4 levels in both factors, versus just two levels each in Baker et. al. The p-values resulting from Baker et. al. ANOVA were approximately of the order of 0.01 to 0.1. As a result, they could use the standard p¡0.05 as a significance threshold. By contrast, our p-values were orders of magnitude less than 0.05. This leads to the second main difference between our studies. Instead of using p-values as a report on selectivity, which would involve arbitrarily selecting a significance threshold with no rationale, we report F values directly. Instead of tabulating the number or proportion of cells exhibiting significant learning effects, we compare F values directly against each other (paired t-tests). This provides a cleaner, more direct report of the selectivity of each cell than Baker et. al.

2.2.7 Analysis: One-Factor Selectivity

To assess selectivity differences between Trained and Complementary responses, we could not use two-way ANOVA because the 8 stimuli of the Trained or Complementary sets do not comprise an exhaustive cross design (not all the levels of each factor are present). So we used three different metrics to assess selectivity. These metrics were applied to Canonical-Protocol cells.

2.2.7.1 One-Way ANOVA

The simplest metric is to use regular one-way ANOVA. This metric tests against the null hypothesis that all eight responses are from the same distribution, and the resulting F value can be interpreted as a measure of general selectivity in the same way that two-way ANOVA was used above.

2.2.7.2 Sparseness

It is established that IT neurons exhibit sparse coding. Sparseness was calculated as the inverse of *response density* [55] across the eight stimuli:

$$RD = \langle \frac{\langle x_i \rangle^2}{\langle x_i^2 \rangle} \rangle_i$$

The resulting value is of range 0 (not sparse) to 1 (sparse). High sparseness indicates a distribution characterized by the presence of one or a few high-responses amongst many low responses. This is in contrast with one-way ANOVA, in which high F values simply point to high selectivity but no further information on the distribution that generates that selectivity.

2.2.7.3 Maximum Interaction Residual

The final metric is the maximum of the interaction residuals from two-way ANOVA. Here we revisit the two-way ANOVA applied over Canonical reponses. two-way

ANOVA models each stimulus response as

$$FR_{Top=t,Bottom=b} = \mu_{Grand} + \mu_{Top=t} + \mu_{Bottom=b} + \mu_{Top=t,Bottom=b} + \varepsilon$$

where the first term is the grand mean, the second and third terms are the main effects, the fourth term is the interaction effect, and the last term is the remaining error. Each Canonical stimulus has a corresponding interaction term. Furthermore, eight interaction terms correspond with Training responses, and eight correspond with Complementary responses. We consider here the maximum of the positive Training interaction terms and the maximum of the Complementary interaction terms, each normalized by the maximum firing rate of the cell.

$$Res_{max} = \frac{max(\mu_{Top=t,Bottom=b})}{max(FR)}$$

This metric is in essence a continuation of the increasing specificity of the three metrics. one-way ANOVA is the most general selectivity metric. Sparseness reports on more specific distributions in which high selectivity is manifest through few high responses. And the Maximum Interaction Residual, by design, reports on only the highest response.

2.2.8 Supplementary Analysis Methods

2.2.8.1 Selectivity Timing

In this section, we endeavor to further examine the ANOVA results by considering the evolving time course of F values. Our goal is to compare the relative timing of main effect contributions (part selectivity) versus interaction effect contribution (part interaction selectivity): does timing for one precede the other or are they similar? To this end, we first smoothed the evoked firing rates of each cell using an asymmetric gaussian kernel as reported in Brincat 2006 [56] (itself adapted from Thompson 1996 [57]). The gaussian kernel was shaped with 15 ms standard deviation on the causal side, 5 ms SD acausal side. This procedure yields a robust estimate of instantaneous response rate that avoids backward bias in time by means of primarily causal weighting. The resulting smoothed firing rates were averaged across all repetitions of each stimulus. The same two-way ANOVA methodology described previously (separate ANOVA performed over canonical and inverted responses) was then applied at every 5ms timepoint. From each cell, six F-value time courses were constructed, corresponding to Top, Bottom, and Interaction values of Canonical and Inverted responses.

For any given set of six F time courses, we used four descriptive parameters to assess the relative timing of main (top and bottom) effects versus interaction effects. Two of the parameters directly describe the time course F values. Time of

Onset F Deviation, referred to as "F_{Dev}", is defined as the time in which F value rose above 25% of the "baseline value". "Baseline value" is the median of the F values during the first 50ms post-stimulus onset. Time of F Peak, referred to as "F_{Peak}", is defined as the time of the maximum F value during the time-course. These parameters are defined for each of the six F time courses. Since our goal is to compare main effects versus interaction effects, we further define "main" and "interaction" versions of the parameters: $F_{Dev,Main}$, $F_{Dev,Int}$, $F_{Peak,Main}$, and $F_{Peak,Int}$. $F_{Dev,Main}$ is defined by taking the minimum F_{Dev} of the four main time courses (Canonical Top, Canonical Bottom, Inverted Top, and Inverted Bottom). $F_{Dev,Int}$ is defined by taking the minimum F_{Dev} of the two interaction time courses (Canonical Interaction, Inverted Interaction). $F_{Peak,Main}$ and $F_{Peak,Int}$ are defined the same way. In this way, we can directly compare main and interaction time points.

We define more timing parameters by first taking the canonical versus inverted ratio of the time-course F values. For example, Canonical Top is divided by Inverted Top at each 5ms interval to create one ratio time course, and Inverted Top is divided by Canonical Top for another ratio (which will be the inverse of the previous ratio). The rationale for considering ratios is that we are not just interested in, for example, a large Canonical Bottom F contribution, but rather a large Canonical Bottom F contribution relative to the Inverted Bottom F contribution. Six time-course ratios were constructed out of the six F values: (Canonical Top / Inverted Top), (Canonical-Bottom / Inverted Bottom), (Canonical Interaction / Inverted Interaction), and the

three inverse ratios (Inverted / Canonical). From these ratios we ascertained when they deviated from unity (which is when Canonical equals Inverted). Time of Ratio F Deviation, referred to as "R_{Dev}", is defined as the time when Ratio F first crossed 1.1. That is, when the numerator F value first deviated to over 10% of the denominator F value. "Time of Significant Ratio F Deviation", referred to as "R_{SigDev}", is defined as the time when Ratio F first deviated above a specified threshold. The threshold is determined by first noting that both F values would come from the same F distribution (either from a main effect with three df in the numerator of the F distribution, from an interaction effect with 9 df in the numerator of the F distribution). The threshold was chosen to be the p = 0.05 threshold of that F distribution. In essence, the threshold is chosen such that if the denominator F value had been 1 (a null result for an F distribution), then the numerator F value is significant (one-tailed) at p=0.05. The threshold is 2.11 and 1.66 for main effect ratios and interaction effect ratios respectively. Similar to the definition of $F_{Dev,Main}$ and others above, we define "main" and "interaction" versions $R_{Dev,Main}$, $R_{Dev,Int}$, $R_{SigDev,Main}$ and $R_{SigDev,Int}$ by taking the minimum R_{Dev}, and R_{SigDev} values across the four main or two interaction values.

Thus, for each individual cell, eight parameters are calculated: $F_{Dev,Main}$, $F_{Dev,Int}$, $F_{Peak,Main}$, and $F_{Peak,Int}$. $F_{Dev,Main}$, $R_{Dev,Main}$, $R_{Dev,Int}$, $R_{SigDev,Main}$ and $R_{SigDev,Int}$. This analysis is conducted over all Full-Protocol cells, giving us a population of values for each parameter. From the population of values, we consider the mean and

standard error.

Furthermore, we also consider aggregate F time courses. Aggregate F time courses are constructed by averaging F time courses across Full-Protocol cells. So a single set of six aggregate F time courses are defined for the entire population of Full-Protocol cells. Ratio time courses are constructed the same way as described above, but from the aggregate F time courses. We then use the same eight timing parameters to characterize the aggregate F and Ratio time courses. The parameters used to describe aggregate time courses are denoted with an "agg": $Fagg_{Dev,Main}$, $Fagg_{Dev,Int}$, $Fagg_{Peak,Main}$, $Fagg_{Peak,Int}$, $Ragg_{Dev,Main}$, $Ragg_{Dev,Int}$, $Ragg_{SigDev,Main}$ and $Ragg_{SigDev,Int}$.

2.2.8.2 Categorical Representation

To assess intra-category versus inter-category variance, we applied Nested ANOVA to the Morphed responses. The ANOVA was structured with two factors, with "category" being the main (top) factor and "morph" being the nested (lower) level. The category factor had either 8 or 16 levels spanning Training, Complementary, Canonical, or Inverted responses. The morph factor had five levels corresponding to the five morphs generated for each category in the Morph protocol. Nested ANOVA is chosen as opposed to two-way ANOVA because the five morphs of one category do not correspond to the five morphs of another category. From the nested ANOVA we compute $F_{Ratio} = F_{Category} \ / \ F_{Morph}$. High F_{Ratio} means that more variance is

captured by category over morphs, indicating a high degree of categorical invariance.

2.2.8.3 Morphed and Task Responses

Finally, we report preliminary data on how Morphed and Task (and Task-Morph) responses differ from Behavioral responses (which are unmorphed and passive). The data shown here was taken over a small subset of cells and does not satisfy requirements for conclusive interpretation.

From the Morph and Task protocols we could show our stimuli under four conditions:

- Passive, Unmorphed stimuli (Behavioral Protocol). Denoted "PassUnmorph".
- Passive, Morphed stimuli (Morphed Protocol). Denoted "PassMorph".
- Active, Unmorphed stimuli (Task Protocol, 2nd stimulus). Denoted "Active-Unmorph"
- Active, Morphed stimuli (Task Protocol, 1st stimulus). Denoted "ActiveMorph".

We made multiple comparisons between the four groups. We compared firing rate, one-way ANOVA and Sparseness. Comparisons between two protocols are made on the set of cells from which the two protocols were executed. This set of cells differs from comparison to comparison. The comparisons between Passive Morphed and Active Morphed, for example, have by far the smallest number of cells as these were

the cells shown the Task-Match protocol. As explained in the protocols section, Task-Match protocols were introduced relatively late and only six cells were recorded with Task-Match.

2.2.8.4 Location Effects

We examined whether there was a correlation between any spatial dimension describing the location of recorded cells and any of the metrics discussed thus far. Specifically, we considered four spatial dimensions: three stereotaxic A-P, M-L, and D-V dimensions as well as the first principal component resulting from principal component analysis of the 3D spatial locations of all the cells. Multiple metrics were assessed by their correlation with these spatial dimensions. The metrics included two-way ANOVA F values, the three Trained-versus Complementary metrics for Aim1-part 2 (one-way ANOVA F, sparseness, normalized maximum interaction residuals), and the 8 timing parameters described for aggregate time courses (see section 2.2.8.1 Selectivity Timing). Each candidate metric was correlated with each spatial dimension across the appropriate cell population (i.e. across 57 Full-Protocol cells for the metric Inverted Top F, but across 77 Canonical-Protocol cells for sparseness). We also considered individual monkeys as well as aggregate data.

2.3 Results

2.3.1 Part 1: Confirmation of Learned Increase in Part Selectivity

2.3.1.1 Part Selectivity

Figure 2.4a and 2.4b show an example response profile of an individual cell to the Behavioral stimuli (see Methods, 2.2.3.3). This animal was trained on set 1 (here outlined in gray). The starred stimulus (maximum evoked response) evoked an average response rate of 75.1 spikes/s. Two-way ANOVA of the Canonical responses yielded Main (top and bottom) and interaction F values of 197.5, 38.8, and 19.6 respectively and corresponding p-values of 1.53e-82, 2.10e-22, and 6.54e-28.The same values for Inverted responses are F values of 29.0, 9.29, and 6.68, with p-values of 7.15e-16, 8.20e-05, and 1.83e-08. Across the population of 57 Full-Protocol cells (see Methods, 2.2.5), top main effects appeared to be equivalent for Canonical and Inverted stimuli (Figure 2.4c, $p = 0.37$, paired t-test, two-tailed. For individual monkeys: $p = 0.86$ for Monkey 1; $p = 0.38$ for Monkey 2), while bottom main effects were stronger for Canonical stimuli (Fig 6d, p = 3.2e-3 overall ; $p = 0.097$ Monkey 1 ; $p = 0.012$ Monkey 2. The smaller number of neurons, 20, in Monkey 1 contributed to the $p > 0.05$, although the trend clearly favored Canonical over Inverted). Thus we reproduce the moderate increase in part selectivity observed by Baker et. al.

2.3.1.2 Part Interaction Selectivity

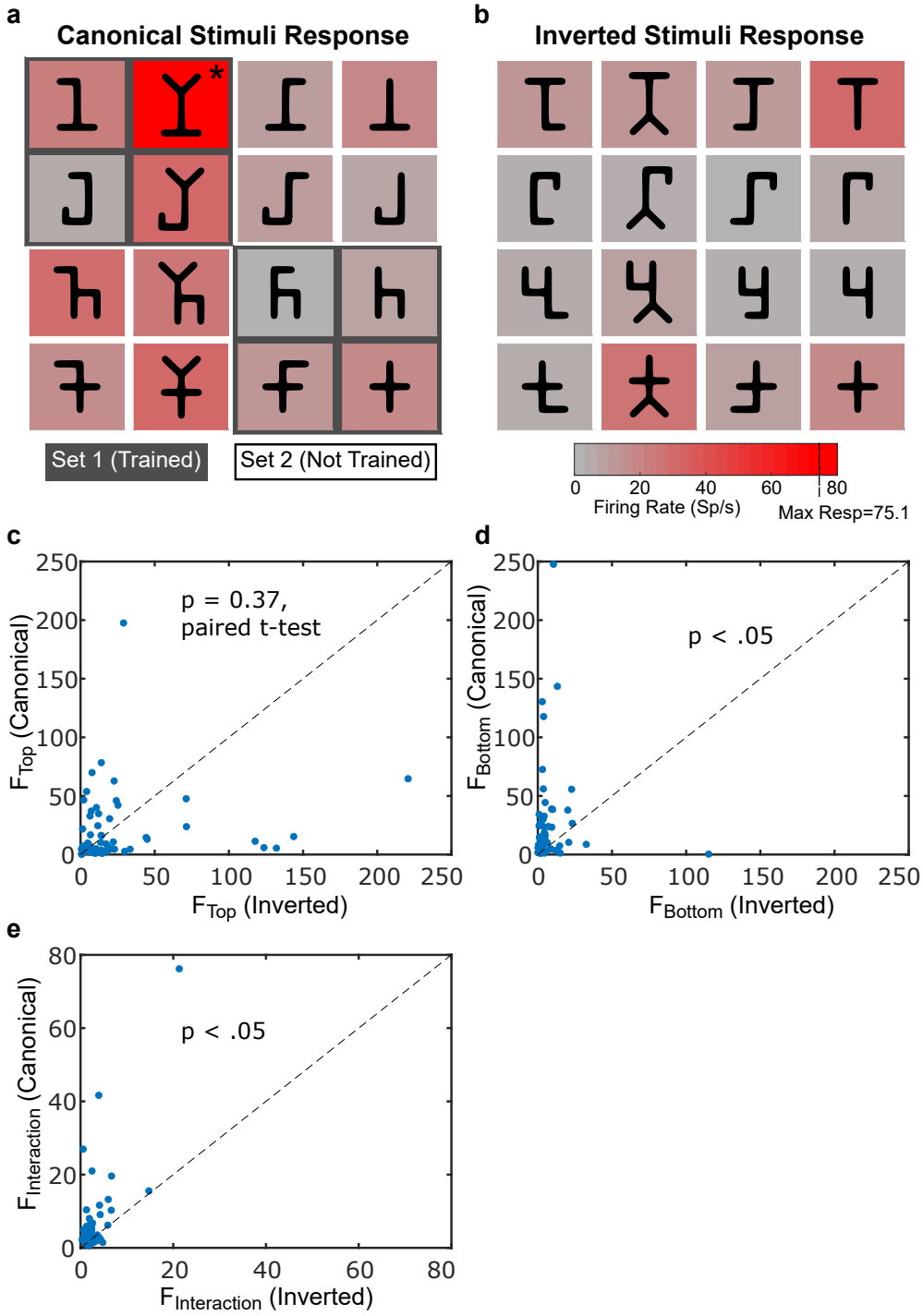
More importantly, we replicated the critical result in Baker et. al. Namely, we see a strong learning effect for part interaction. Interaction effects were stronger over the population of Canonical responses as compared to Inverted responses (Figure 2.4e; p = 2.3e-3 overall ; p = 0.0814 Monkey 1 ; p = 0.0142 Monkey 2. Again, a small number of Full Protocol cells from Monkey 1 contributes to p > 0.05).

Figure 2.4: Aim1 Part 1: Canonical Versus Inverted ANOVA Results

(a,b) An example response profile of an individual cell to Behavioral stimuli. The gray outlines signal the Trained stimuli for this monkey. The background color indicates the relative evoked response for each corresponding stimuli. The "redder", the higher the evoked potential. The stimulus evoking the maximum firing rate (75.1 sp/s) is starred. This example cell is more selective for Canonical stimuli than Inverted, (c-e) Two-way ANOVA was performed for each cell's response both to Canonical and Inverted stimuli (see Methods). These scatterplots compare the resulting Canonical versus Inverted F values for Top (c), Bottom (d), and Interaction (e) effects. Paired t-tests were used to compare the populations of Canonical versus Inverted F values. The resulting p-values are displayed in the corresponding graphs. Top F values were not significantly different across Canonical and Inverted values. However, Canonical bottom and interaction F values were significantly different than their respective Inverted values.

(next page)

Figure 2.4: Aim1 Part 1: Canonical Versus Inverted ANOVA Results



2.3.2 Part 2: Counterevidence to Learned Increase in Holistic Selectivity

The results so far echo the findings of Baker et. al. in that they suggest learned selectivity for whole, familiar objects. However, given our extended stimulus design which includes the Complementary set of categories, we could test whether increased selectivity was specific to familiar objects, or more generally observable for familiar parts, even when combined into unfamiliar objects. Figure 2.5a and 2.5b showcase an example cell that exhibits a similar learning effect to the cell displayed in Figure 2.4. There is a learning effect apparent in comparing Canonical versus Inverted selectivity, and specifically, an increase in part interaction selectivity (over simply part selectivity) is exhibited. However, a major difference from the cell in Figure 2.4 is that for this cell, the enhanced selectivity highlights Complementary stimuli instead of the Trained stimuli. The starred stimulus was never shown during training, but elicited the highest response and resulted specifically from a highly significant interaction effect ($F = 21.0$, $p = 1.67e\text{-}25$, interaction residual for the starred stimulus = 12.8 spikes/s).

To test the generality of this result, we turn to the 77 Canonical-Protocol cells using metrics of one-way ANOVA, sparseness, and normalized maximum interaction residuals to compare between Trained and Complementary selectivity (see Methods, 2.2.7).

Figure 2.5c-e shows that for all three metrics we do not see a significant difference

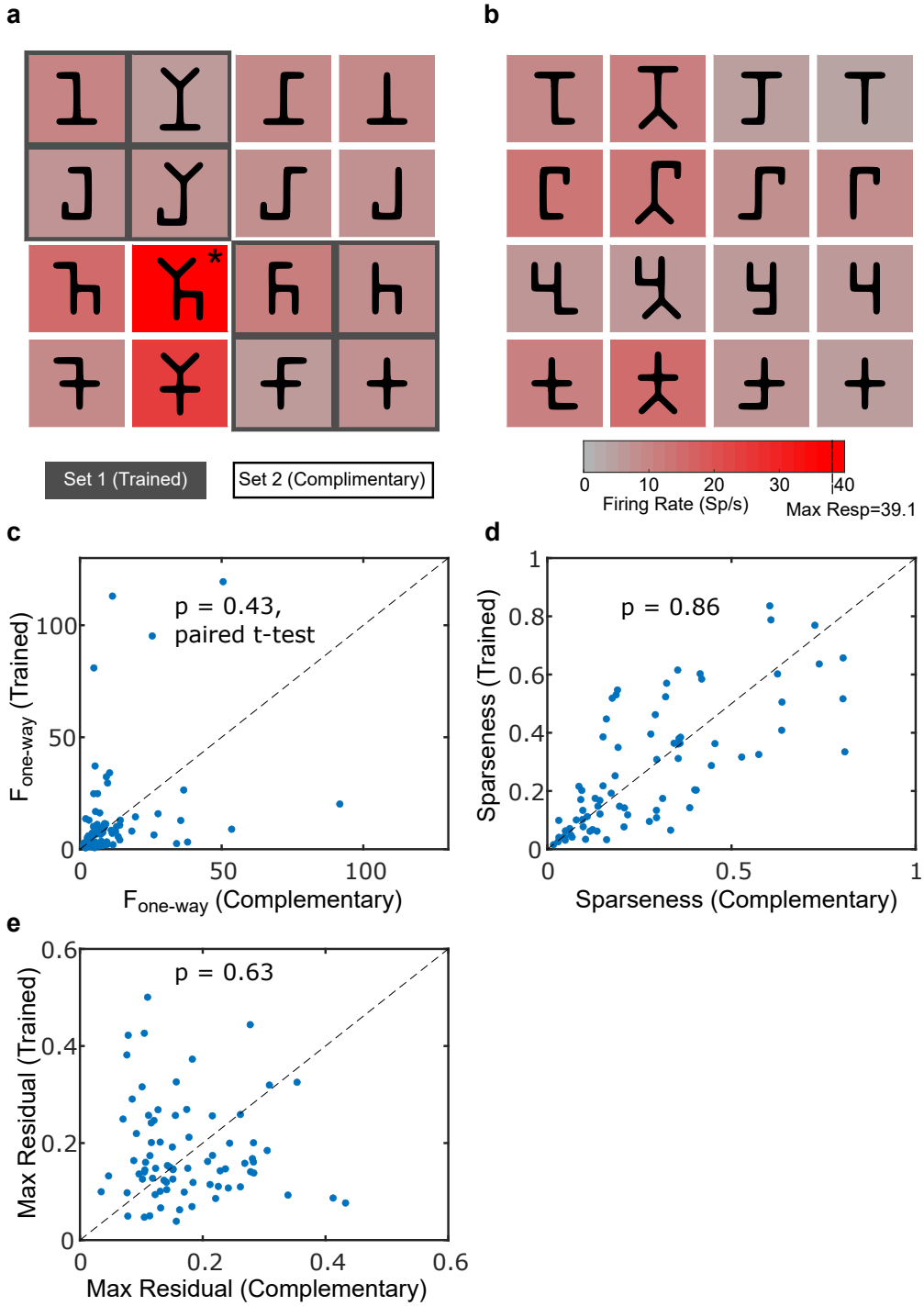
Figure 2.5: Aim1 Part 2: Trained Versus Untrained Results

(a,b) example response profile of an individual cell to Behavioral stimuli. The gray outlines signal the Trained stimuli for this monkey. The background color indicates the relative evoked response for each corresponding stimuli. The stimulus evoking the maximum firing rate (39.1 sp/s) is starred. This stimulus also exhibits the highest supra-additive interaction residual at 12.8 sp/s. Importantly, this cell exhibits a learning effect similar to the cell displayed in Figure 6: increased selectivity across Canonical versus Inverted stimuli, and enhanced part interaction selectivity. However, in this cell, the increased part interaction selectivity highlights Complementary stimuli (the starred stimulus is not a Trained stimuli, but Complementary instead). This phenomenon is not readily explained by whole-object selectivity. (c-e) Three different metrics are used to compare Trained versus Complementary responses over the population of Canonical-Protocol cells (see Methods): c) one-way ANOVA. d) Sparseness, e) normalized maximum interaction residuals. For all three metrics we do not see a significant difference between Trained and Complementary selectivity.

(next page)

between Trained and Complementary selectivity. That means we could not see general selectivity differences (one-way ANOVA), response profiles of familiar objects were not sparser than unfamiliar objects, and the supra-additive residuals were not significantly stronger for Trained versus Complementary objects.

Figure 2.5: Aim1 Part 2: Trained Versus Untrained Results



2.3.3 Supplementary Results

2.3.3.1 Selectivity Timing

In this section, we compare the relative timing of main effect contributions versus interaction effect contribution to assess whether one precedes the other. We expanded on the ANOVA result by computing smoothed firing rate and then computing F time courses, aligned to stimulus onset. From the F time courses, we calculated eight descriptive parameters, $F_{Dev,Main}$, $F_{Dev,Int}$, $F_{Peak,Main}$, $F_{Peak,Int}$, $R_{Dev,Main}$, $R_{Dev,Int}$, $R_{SigDev,Main}$ and $R_{SigDev,Int}$ to describe assess the relative timing between main effects and interaction effects. These parameters were calculated for both individual cell responses, and on the aggregate F time courses, calculated by taking mean F time courses across all Full-Protocol Cells (i.e. $Fagg_{Dev,Main}$). See Methods (2.2.8.1) for a full explanation of the parameters and how they were calculated

Figure 2.6a shows the six aggregate F time courses. The same trends of Canonical and Inverted F values from part 1 results are apparent here. Canonical Bottom and Interaction time courses deviate above Inverted Bottom and Interaction time courses. Inverted Top time course trends higher than Canonical Top. However this is primarily due to a few outlier cells skewing the mean (not shown), and the level of deviation is not as high as the bottom or interaction effects. Aggregate parameters ($Fagg_{Dev,Main}$, $Fagg_{Dev,Int}$, $Fagg_{Peak,Main}$, and $Fagg_{Peak,Int}$) are denoted in the single dots and x's (black and red colors). The population of individual parameters are visualized with

error bars next to the corresponding aggregate dots or x's. The error bars are centered about the sample mean and the span of the bars denote the standard error of that particular parameter across Full-Protocol cells.

Figure 2.6b shows aggregate Ratio time courses. For simplicity, three of the six ratios are shown. The other three are simply inverses of the three shown. In the same fashion as described above, the aggregate parameters are displayed as dots and x's and the population of individual-cell parameters are visualized with the error bars.

Tables 2.2 and 2.3 quantify the parameters visualized in Figure 2.6. Table 2.2 displays the relative occurrences of the eight descriptive metrics (with corresponding main-versus-interaction pairs displayed by row). Four scenarios are considered (by column). 1) There could have been a main effect and no interactive effect. 2) Both main and interaction effect was present with main effect preceding interaction effect. 3) Both effects are present with interaction preceding main. 4) Interaction effect present but no main effect. A majority of cells across metrics show the main effect preceding an interaction effect, a minority show the interaction effect preceding the main effect, and no cells exhibit one effect without the other.

Table 2.3 quantifies statistical descriptors (mean and standard error) of the individual-cell effects, as well as aggregate parameter values in parenthesis. Again using the population of individual-cell parameter values, the comparison of main-versus interaction pairs (for example, $F_{Dev,Main}$ versus $F_{Dev,Int}$) show significant positive mean shift time (shift = interaction time - main time) across all metric pairs (Table 2.3,

column 4 and 5).

We find mean shift times ranging from 26.5 to 47.1 ms depending on parameter considered. All mean shift times were significant. In conclusion, we looked at a variety of timing parameters to characterize F time courses and find clear evidence that main effects precede interaction effects by at roughly 30ms.

Table 2.2: Counts & Proportions of the Relative Timing of Main Versus Interaction Selectivity.

	Main Effect, No Int Effect	**Main precedes Int Effect**	**Int precedes Main Effect**	**No Main Effect, Int Effect**
	count (*percentage*)			
Onset F Deviation ($F_{Dev,Main}$ & $F_{Dev,Int}$)	0 (0.0%)	44 (77.2%)	13 (22.8%)	0 (0.0%)
Onset Ratio Deviation ($R_{Dev,Main}$ & $R_{Dev,Int}$)	4 (7.0%)	40 (70.2%)	13 (22.8%)	0 (0.0%)
Onset Significant Ratio Deviation ($R_{SigDev,Main}$ & $R_{SigDev,Int}$)	4 (7.0%)	41 (71.9%)	12 (21.1%)	0 (0.0%)
Peak F Deviation ($F_{Peak,Main}$ & $F_{Peak,Int}$)	0 (0.0%)	37 (64.9%)	20 (35.1%)	0 (0.0%)

Table 2.3: Statistical Descriptors of the Relative Timing of Main Versus Interaction Selectivity.

	Main Time, (ms)	**Int Time** (ms)	**Int - Main Time Diff** (ms)	**Int vs Main** Paired t-test p
column-specific legend	*indiv mean* ± *indiv std err* (*aggregate*)			*indiv only*
Onset F Deviation ($F_{Dev,Main}$ & $F_{Dev,Int}$)	59.0 ± 2.9 (60)	89.7 ± 5.5 (75)	30.7 ± 5.5 (15)	6.26e-07
Onset Ratio Deviation ($R_{Dev,Main}$ & $R_{Dev,Int}$)	68.8 ± 2.3 (65)	109.1 ± 6.6 (75)	41.3 ± 7.0 (10)	5.70e-7
Onset Significant Ratio Deviation ($R_{SigDev,Main}$ & $R_{SigDev,Int}$)	87.0 ± 2.4 (85)	123.1 ± 6.4 (140)	37.1 ± 6.8 (55)	2.47e-6
Peak F Deviation ($F_{Peak,Main}$ & $F_{Peak,Int}$)	121.0 ± 4.3 (130)	147.5 ± 5.2 (155)	26.5 ± 5.5 (25)	1.24e-5

2.3.3.2 Categorical Representation

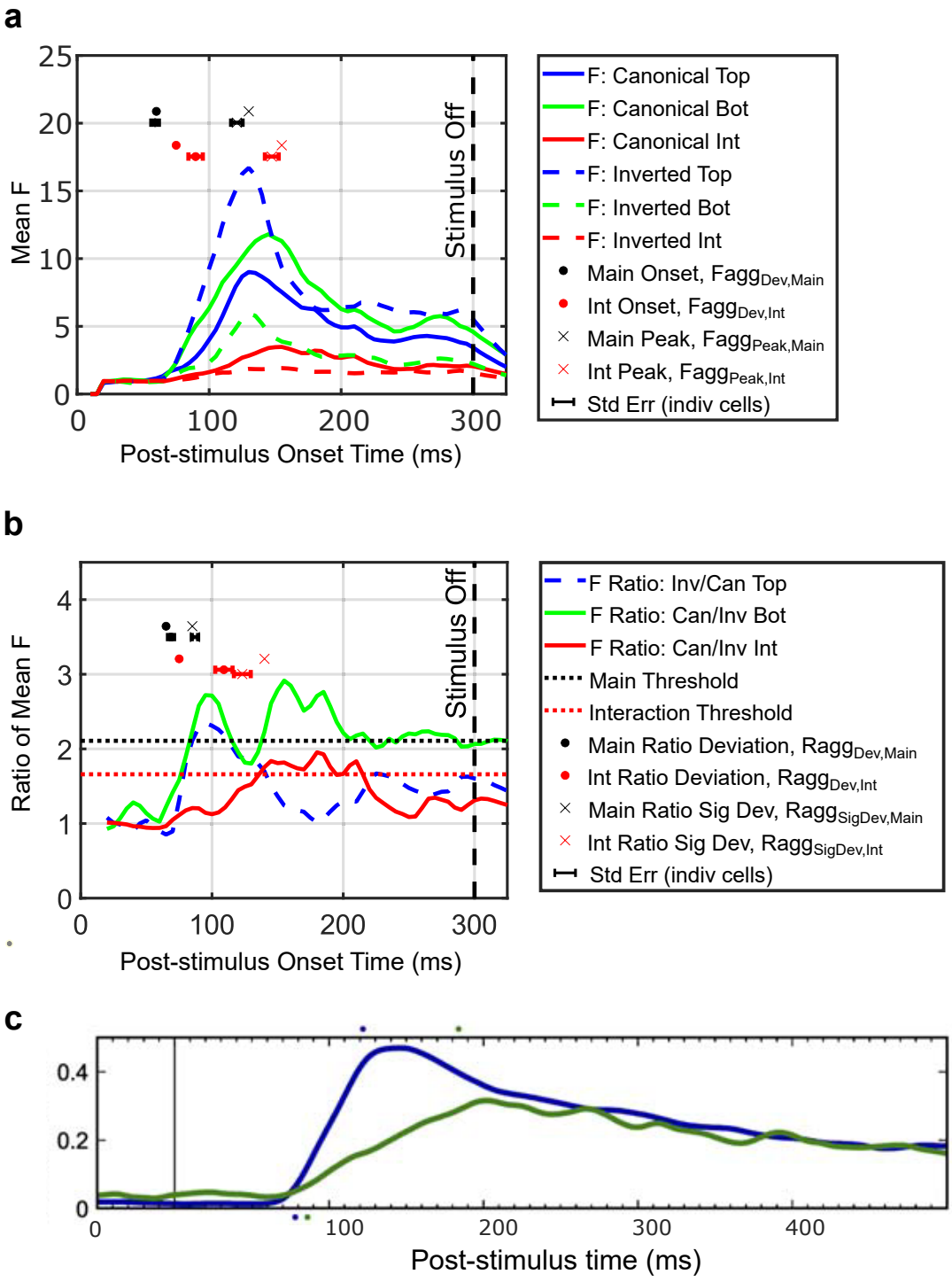
We tested compared intra- versus inter category variance by applying nested ANOVA over the Morphed responses (see Methods, 2.2.8.2). Figure 2.7 compares

Figure 2.6: Selectivity Timing

Evoked firing rates of each cell were smoothed, and two-way ANOVA was performed on each cell at every 5ms interval. From the resulting individual-cell F time courses, we averaged over all the cells to produce ""aggregate F time courses"". The traces shown in (a) are the aggregate time courses. Note that Bottom and Interaction F time courses exhibit the learning effect (Canonical ¿ Inverted) reported in part 1 result (two-way ANOVA, no time-course). We also define F Ratio time courses, done by dividing a canonical F time course at each time point with the corresponding inverted F time course. Once again, individual-cell and aggregate versions of ratio time courses are defined. The traces shown in (b) are three of the six aggregate time courses (the other three are simply the inverse of the shown three). The Ratio time courses are compared to main and interaction thresholds (horizontal dotted lines) to determine time points of significant deviation of canonical and inverted F (see methods). We defined eight descriptive parameters (see methods) to quantify the timing and comparison of canonical versus interaction time courses. Four of the parameters are defined for F traces and four are defined for Ratio traces. Each of the parameters were defined over both aggregate and individual-cell traces. The aggregate versions are shown in both plots (a) and (b) as single dots or x's. The population of individual-cell values (across the Full-Protocol cells) are visualized by the corresponding error bars next to the aggregate dots or x's. The error bars are centered around and span the mean and standard error of the population of values respectively. The plots (a) and (b) are meant to be read along with Tables 2 and 3. The conclusion drawn from these graphs is that main effects precede interaction effects by about 30-40 ms. This bears striking resemblance to the results of Brincat 2006 [56]. The plot in (c) is a reprinted graph from that study, in which experimenters report differential timing effects between linear and nonlinear representations of stimuli in posterior IT cortex. They modeled neuronal responses to 2D shape-contour stimuli with linear and nonlinear components and considered the relative contributions linear and nonlinear components have over the time course of smoothed firing rates. They report significant linear contributions preceding significant nonlinear contributions (indicated by the stars above the graph) by about 60ms. Loosely drawing parallels between linear/nonlinear versus main/interaction, results of this project closely mirrors the findings of Brincat 2006 and suggests that object recognition happens in stages along different timescales.

(next page)

Figure 2.6: Selectivity Timing



Trained versus Complementary F_{Ratio} as well as Canonical versus Inverted F_{Ratio}. Both comparisons fail to demonstrate any significant difference. The maximum Trained and Canonical F_{Ratio} were higher than the maximum Complementary and Inverted F_{Ratio} respectively. Also, the population of Trained FRatio appears to be trending higher than Complementary F_{Ratio} with twice as many cells above versus below the unity line. However, with the small amount of cells in which Morph protocol was recorded (18), the data is inconclusive.

2.3.3.3 Active Versus Passive Context

Figure 2.8a shows a schematic of the four contexts that we show our stimuli (PassUnmorph, PassMorph, ActiveUnmorph, and ActiveMorph, see Methods, 2.2.8.3) and the protocols they come from. We also highlight the four comparisons we make amongst the contexts, with each four dotted rectangles representing a comparison. The rectangles are in the same relative spatial location as each of the four comparison scatterplots in 10b-e. The numbers in each of the rectangles denote the number of cells used for the comparison. The numbers differ from each other because of the differential amount of cells each protocol was shown to.

Figure 2.8b shows firing rate comparisons amongst the four contexts. Blue dots represent comparisons of individual stimuli. Displayed morphed responses are the mean elicited response over all morph stimuli of that protocol (5 for passive Morphed protocol, 3 for active Task protocol). Red dots represent comparisons of mean re-

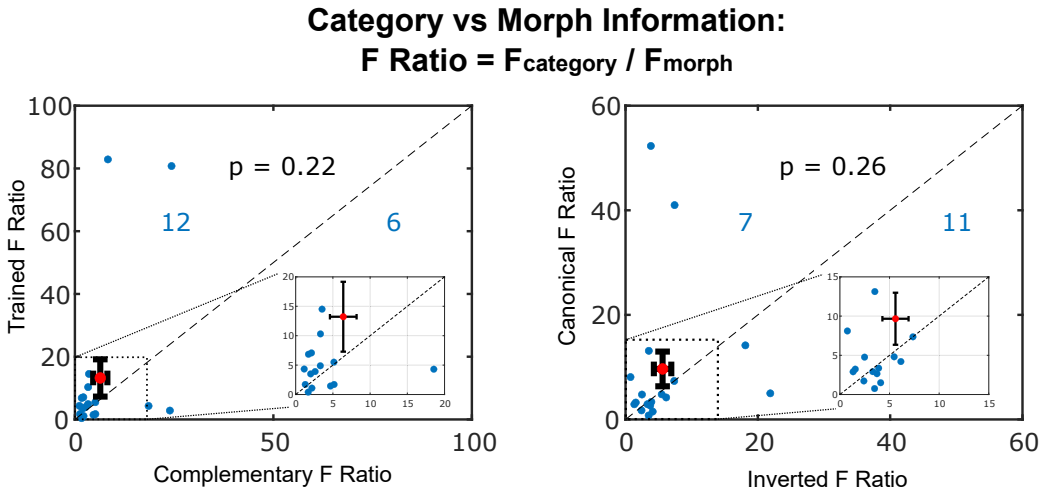


Figure 2.7: Nested ANOVA results applied over Morphed responses. This analysis compares intra- versus inter- category variance, where intra category variance refers to variance explained by different morphs of the same categories, and inter-category variance refers to the variance explained across by different categories. The data plotted is FRatio, which is inter-category variance divided by intra-category variance. The higher the FRatio, the more the response profile was influenced by various categories and the less by various morph levels. The two graphs show Trained versus Complementary and Canonical versus Inverted FRatios respectively (see Methods). Both comparisons fail to demonstrate any significant difference. Note that visually, Trained and Canonical FRatios may be trending higher than the respective Complementary and Inverted FRatios (compare the maximum Trained/Canonical values versus the maximum Complementary/Inverted values. Also, twice as many Trained-vs-Complementary data points are above the unity line than vice versa). However, the number of total cells (18) may be too small to be conclusive.

sponse across all stimuli of their respective protocol and cell. Numbers indicate the number of comparisons above and below the unity line. For the firing rate graphs, only numbers of mean response comparisons are shown for simplicity.

PassUnmorph and PassMorph exhibit a fairly tight linear relationship, with PassMorph trending higher than PassUnmorph (paired ttest, p=9.12e-32). By comparison, PassUnmorph and ActUnmorph have a looser relationship. There appears to be two populations of cells, with some exhibiting a tight, fairly equal relationship while in others, ActUnmorph responses are significantly larger than PassUnmorph with an almost multiplicative gain. ActUnmorph and ActMorph have the tightest linear relationship and are not significantly different. PassMorph and ActMorph (even though there are only six cells, far less than other relationships) exhibit a similar multimodal relationship as PassUnmorph vs ActUnmorph, with most cells exibiting a tight, equal relationship and one outlier cell showing a large increase in ActUnmorph responses.

Before trying to draw conclusions from the trend above, it is worth noting again that there are too few cells to draw lasting conclusions and this data is preliminary. The above results can be recast as the following: PassUnmorph responses are fairly close to (but slightly smaller than) PassMorph responses. ActUnmorph and ActMorph responses (which are from the same protocol) can be broken into roughly two categories, with some being equal to PassUnmorph and PassMorph responses, and some being markedly greater. This difference could indicate a passive versus active signal modifying the response in some cells.

Figure 2.8c also shows firing rate, but only amongst (and color coded by) the six cells of the Task-Match protocols. The same trends from above apply, but with a more focus on a small subset. The tight relationship between PassUnmorph and PassMorph as well as the tight relationship between ActiveUnmorph and ActiveMorph still is apparent. Relationships between Active and Passive firing rates also exhibit linear relationships, except for one outlier cell in which Active responses seem to far outpace Passive responses.

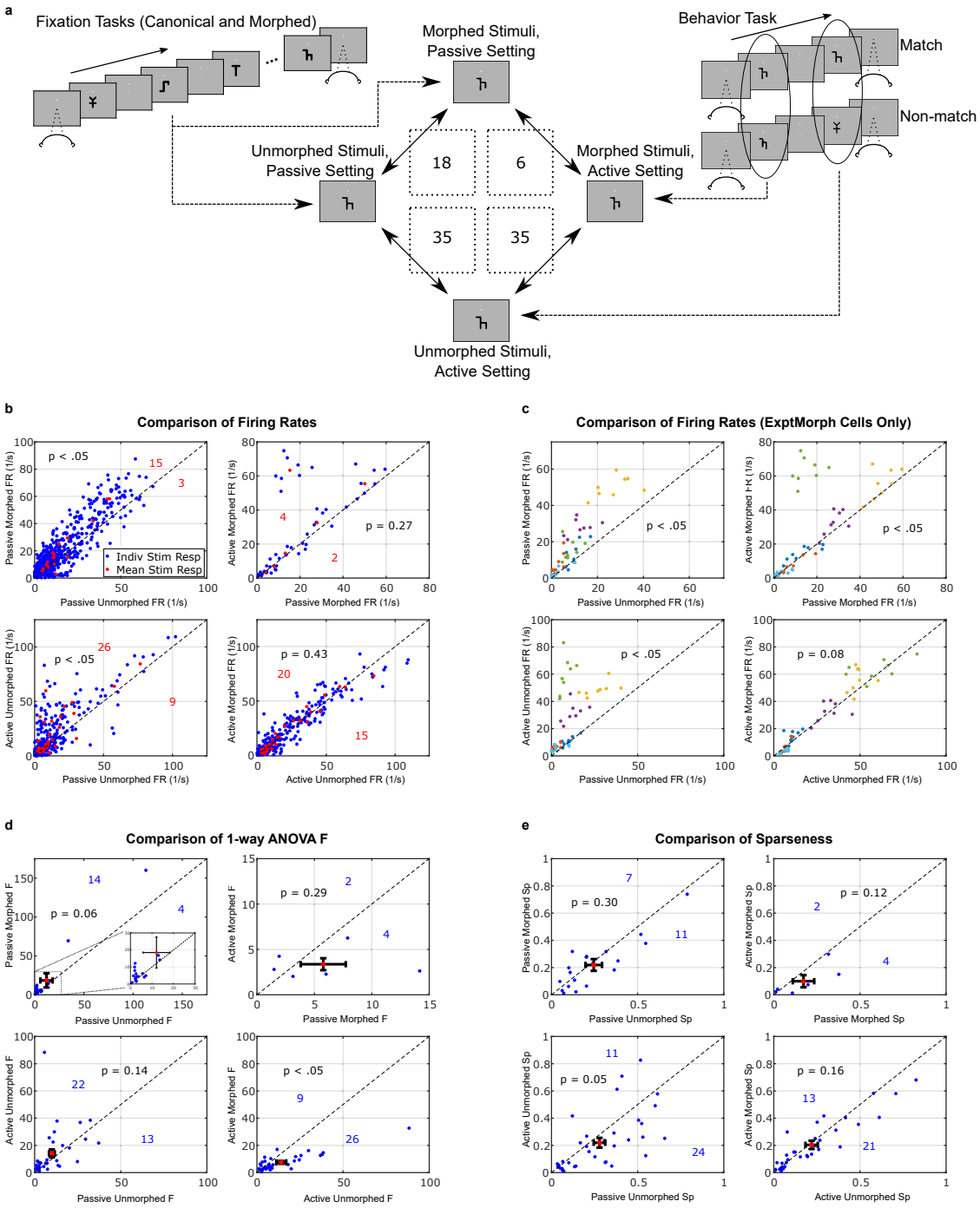
Figure 2.8d and 2.8e show one-way ANOVA F and sparseness comparisons between groups. It is worth noting that both of these metrics are unaffected by possible gain changes in firing rate that was observed above. Sparseness in particular is not affected by definition. ANOVA is unaffected as long as the mean squared error is multiplied by the same weight as non-error mean squared error. Across all comparisons of these metrics, all are not significant, and visual inspection of the graphs do not reveal anything striking with the lone exception being the one-way ANOVA comparison of ActUnmorph to ActMorph. Here, ActUnmorph responses show a striking linear relationship with, but also are significantly greater than responses of ActMorph. ActMorph and ActUnmorph respectively comprise the stimuli of the first (sample) and second (match) stimuli of the trials in the Task protocol. Thus, the greater F values of the ActUnmorph responses can be interpreted as greater F values associated with the match period over the sample period.

Figure 2.8: Active-vs-Passive and Morph-vs-Unmorphed Effects

(a) A schematic of the four contexts through which that our stimuli are present, the protocols that they come from, and the relationships between them that we show in parts (b-e). In the central diamond shape in the schematic, the upper node represents Morphed stimuli responses in a passive context and comes from the Morph Protocol. The left node represents Unmorphed Stimuli responses in a passive context and comes from the Behavioral Protocol. The right and bottom nodes, both coming from the Active Task Protocol, represent Morphed and Unmorphed stimuli responses respectively. The four dotted rectangles represent the comparison between the two contexts they lie in between. The rectangles are in the same relative spatial location as the four comparison scatterplots shown in (b-e). The numbers in each of the rectangles denote the number of cells used for the comparison. The numbers differ from each other because of the differential amounts of cells each protocol was shown to. (b) Firing rate comparisons among the four contexts. Blue dots represent comparisons of individual stimuli, red dots represent comparisons of mean response across all stimuli of their respective protocol and cell. Numbers indicate the number of comparisons above and below the unity line. PassUnmorph and PassMorph exhibit a tight unity relationship as does ActiveUnmorph and ActiveMorph. When comparing active contexts to passive contexts, there appears to be two populations of cells, with one (larger) population exhibiting a tight unity relationship, but another population of cells exhibiting enhanced responses to the active contexts. c) Also firing rate comparisons, but with a focus on the subset of six cells in which all protocols including "Task-Match" were shown. The individual responses are now color coded by cell. The same relationships described in (a) apply here. Note that five of the six cells exhibit unity among passive and active contexts while one outlier cell exhibits a large increase in the active responses over passive responses. (d,e) Comparisons of one-way ANOVA F and sparseness respectively. Both metrics are unaffected by possible gain changes in firing rate as observed above. All the comparisons do not reveal significant differences with the exception of the one-way ANOVA comparison of Active Morph F and Active Unmorph F. Active Unmorph responses show a linear but significantly greater value than Active Morph. Active Unmorph is the second of the two stimuli shown in the Task protocol. Thus, the greater F values of Active Unmorph could be interpreted as greater selectivity the match period (second stimulus presentation) in a match-to-sample context.

(next page)

Figure 2.8: Active-vs-Passive and Morph-vs-Unmorphed Effects



2.3.3.4 Location Effects

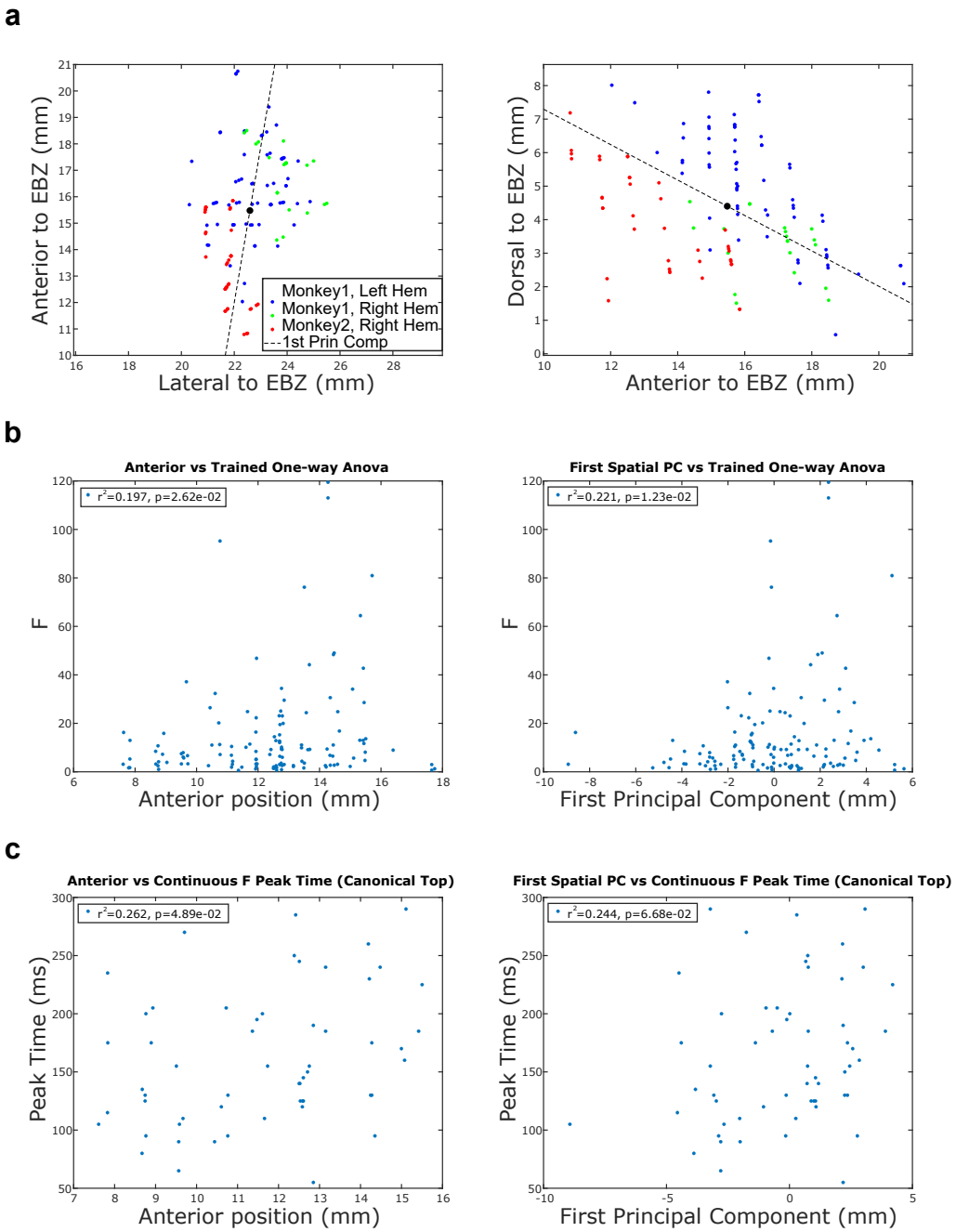
We found no trends to note. The two graphs of Figure 2.9a show the locations of recordings, color coded by recording sessions (three separate sessions: monkey 1 left hemisphere, monkey 2 right hemisphere, monkey 1 right hemisphere). As described in Methods, we considered possible relationships between a multitude of metrics with spatial dimensions. We also considered relationships between all cells as well as individual monkeys. In all, 168 possible relationships were considered. The graphs in Figure 2.9b and 2.9c show some example relationships with their correlations. Displayed are one-way ANOVA and F_{Peak} versus AP and PC1 dimensions. Despite the small p-values associated with the correlations (around 0.05), visual inspection shows a very weak trend at best. These were among the best correlations found across all tested metrics. Thus, we conclude that there is no trend with regards to spatial location.

Figure 2.9: Location Effects

(a) The locations of recordings, color coded by recording sessions. (b,c) Example correlations between metrics and spatial location: (b) one-way ANOVA and (c) FPeak. For both metrics, the graphs depict the relationship between the metric and the Anterior-Posterior dimension as well as the first spatial principal component dimension. These metrics were picked because they exhibited the highest correlations across all 168 tested metric-location pairs. However, visual inspection shows a weak trend at best. Thus, we conclude there is no significant trend of any of the metrics with regards to spatial location.

(next page)

Figure 2.9: Location Effects



Chapter 3

Analysis of Shape Tuning Changes Underlying Learned Visual Recognition

3.1 Motivation

Whereas the last chapter essentially explored the same question as Baker et. al. (albeit with new and supplemental areas), this chapter represents new territory. Here we endeavor to explain neuronal shape tuning that results from learning. Whereas the last chapter took a top-down approach, viewing learned neuron responses through the lens of the same stimuli and stimulus parts defined in the experimental design, this chapter's approach is bottom-up. We investigate the shape space of each neuron

independent from and agnostic to learned behavioral shapes from Chapter 2. The ultimate goal is for insights gleaned from this effort to converge back onto Chapter 2 and explain the learned selectivity that we observe.

To constrain a mathematical model, we need far more data points than the response to 32 Behavioral shapes (from Chapter 2). This chapter details the algorithm used to automatically explore the shape space of neurons and provide us with enough data needed to model it. We then describe the particular model used and the reasons for using it. Finally, we detail our attempts to use the model to describe or extend the results from Chapter 2.

3.2 Methods

3.2.1 Genetic Algorithm

In previous studies, we have characterized complex shape tuning with linear/nonlinear models fitted using a genetic search algorithm [56, 58–60]. This version of the genetic algorithm consists of two dimensional letter-like stimuli defined by medial axis topology, similar to the Behavioral stimuli.

3.2.1.1 Genetic Algorithm: Stimulus Generation

Refer to section 2.2.2 (Aim 1 Methods) for an overview of how our stimuli are generated. In the Behavior and Morphed protocols of Aim 1, the medial axis skeletons that define each stimulus were defined and repeated for each stimulus generation. For the genetic algorithm, however, no such regularly defined skeletons are present. The skeletons themselves are randomly generated (or morphed). A randomly generated skeleton will start with a single limb precursor set to a random angle and length. From there, a random boolean decides whether to add another limb or stop the skeleton generation. An added limb will be added to a randomly picked node from the skeleton. Conditions of a maximum of six limbs per skeleton and a maximum of four limbs joined at a single node are enforced. Once the randomly generated skeleton is finished, a surface contour is generated in the same way as outlined in Aim 1 (with widths and smoothness also randomly generated). Consistency is checked: if a smooth and closed surface contour cannot be generated (for example, if limbs are overlapping or crossing), then the stimulus is discarded and the process is repeated.

Morphed versions of the genetic algorithm stimuli are executed in a similar fashion to the generation of Morphed (behavioral) stimuli in Aim 1. Starting from the original skeleton, certain parameters of the stimulus generation are changed, consistency is checked, and the new surface contour is generated. However certain major differences apply. First, Morphed GA stimuli can undergo new morph types that were not allowed in Aim 1 morphing. Limbs can be added or subtracted now. Also, junction

angles are allowed to change. Another major change regards how multiple morph dimensions are handled. In Aim 1, all the allowed morphs (regarding limb length, node width, and smoothness) happened simultaneously to all appropriate elements. That is, all limbs underwent length changes while all nodes underwent width changes simultaneously. In a genetic algorithm morph, only one single morph type is selected and executed from the pool of all possible morph types. The possible morph types are limb addition, limb subtraction, angle changes, limb lengths, node widths, and node smoothness. Furthermore, once a morph type is selected, it is then applied to only a single element. For example, a limb length morph is applied to only a single limb chosen at random instead of all the limbs (as was done in Aim 1).

3.2.1.2 Genetic Algorithm: Protocol

The first stimulus generation contains only randomly generated two dimensional stimuli. Evoked responses (averaged across 5 repetitions) were ranked into 10 bins with equal numbers of stimuli. In the second generation, 10–20% of stimuli were randomly generated. The rest were morphed descendants of ancestor stimuli from the first generation, selected randomly in equal numbers from the 10 bins. Thus, a typical second generation would contain 10 stimuli generated de novo, 4 descendants of stimuli in the highest response bin, 4 descendants from the second highest bin, etc. In subsequent generations, ancestor stimuli were pooled across all preceding generations and re-binned. Figure 3.1a shows an example cell's response to the genetic

algorithm. The first and second generation show modest evoked responses while in the last generation, the best stimuli evoke relatively high responses.

A drawback of this approach is the large number of free parameters required to quantify complex shape and the consequent dangers of overfitting and instability. We address this by employing two separate lineages in the genetic algorithm. Each lineage is independent of the other. Stimuli selected for morphs always stayed within each lineage with no "contamination" between the two. Each generation contained 40 stimuli per lineage, for a total of 80 stimuli. Figure 3.1b shows the final generation of the two lineages of an example cell. Note that both lineages independently converged on a similar shape, demonstrating the efficacy of this method.

Figure 3.1: Aim 2 Methods: Genetic Algorithm

a) Example response profiles of the genetic algorithm for one example cell. The red background indicates relative evoked responses. The first generation contains only randomly generated stimuli and evokes only modest or low responses. The second generation involves a mix of new stimuli and morphed stimuli from randomly selected parents from the first generation. This also produces modest evoked responses. By the last generation (for this cell, generation 10), however, the top evoked firing rates are much higher as the algorithm converges on areas of the shape space that drives the cell. b) Separate lineages of the genetic algorithm. The genetic algorithm is executed with two separate lineages in parallel and simultaneously. Each lineage generates and morphs its own stimuli with no cross talk between them. This is done to avoid convergence on local minima. For this example cell, note that the top stimuli of both lineages have converged to the same general shape characteristics, demonstrating the efficacy of this method.

(next page)

Figure 3.1: Aim 2 Methods: Genetic Algorithm

3.2.2 Description of the Medial Axis Model

The purpose of the genetic algorithm is to provide the data needed to mathematically describe a cells' shape space. Following is a description of our efforts to construct such a model.

3.2.2.1 Medial Axis Elements

We attempt to explain neuron behavior by constructing a model describing neural response to the medial axis elements of the shown stimuli. The model is henceforth referred to as the Medial Axis Model. Medial axis representation has been utilized in object recognition in the realms of theory [61–63] and computer vision [64–66]. In the ventral pathway, previous evidence shows that IT neurons represent medial axis structures (along with surface shape) in a parts-based manner [60].

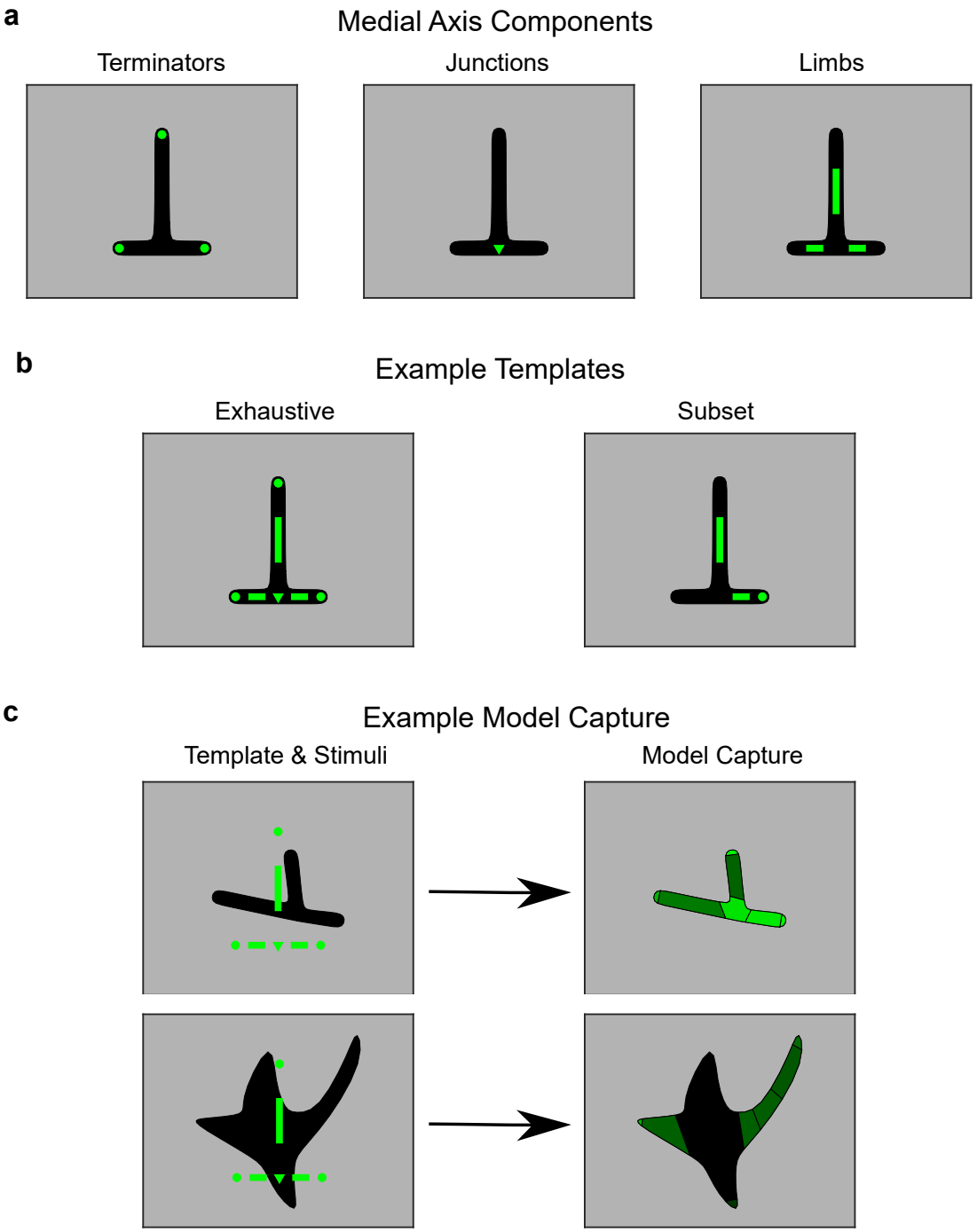
We considered three medial axis elements in our analysis: terminators, junctions, and limbs. Figure 3.2a shows an example stimulus with each of the three types highlighted. Limbs are the single "lines" comprising the skeleton of each letter. Terminators are the terminating endpoint of a single limb. Junctions are the joints conjoining two limbs. These medial axis elements directly follow from the way the stimuli are generated (see section 2.2.2 Stimulus Generation). All the green highlights across Figure 3.2 are conceptual. They were never displayed as part of the actual stimulus.

Figure 3.2: Aim2 Methods: Medial Axis Elements and Template

a) Examples of the three medial axis types considered in the Medial Axis Model. All three graphs are of the same example stimulus. Terminators, Junctions, and Limbs respectively are highlighted. Limbs are the single "lines" comprising the skeleton of each letter. Terminators are the terminating endpoint of a single limb. Junctions are the joints conjoining two limbs. All the green highlights throughout the figure are conceptual. They were never part of the actual stimulus displayed to the monkey. b) Examples of templates. Any template is a nonzero set of medial axis elements that come from a single parent stimulus. The same stimulus from (a) is shown along with two templates drawn from it. The first graph shows a template which contains all medial axis elements from the parent stimulus. The second graph shows a template that contains an arbitrary subset of medial axis elements from the parent stimulus. Regardless of how many elements comprise the template, the template stands apart conceptually from the parent stimuli. c) Examples of comparisons between template and stimuli, called "capture". The exhaustive template from (b) is compared to two stimuli. Each element from the template is matched to an element to the comparison stimulus. The exact pairing of elements result from an exhaustive combinatorial search pairing the same-type elements of the template and stimulus. Each element pair is then compared by a gaussian kernel over dimensions specific to each element type. See Table 3.1 for a summary of the comparison dimensions of each element type. The resulting "individual capture" values are visualized here. The comparison stimuli were segmented into regions corresponding to each medial axis element. The elements on the comparison stimuli that were captured are highlighted according to the individual capture value which range from 0 to 1 (with increasing degree of similarity) and is mapped onto a black-to-green color scale. The overall "capture value" is obtained by averaging the individual captures. Finally, the capture values are inputted into a sigmoid function to obtain the predicted firing rate of the associated stimulus.

(next page)

Figure 3.2: Aim2 Methods: Medial Axis Elements and Template



3.2.2.2 Medial Axis Templates

Each stimulus can be defined by the set of medial axis elements that comprise it. The first graph of Figure 3.2b shows an example stimulus and all of its medial axis elements. We now define the concept of a medial axis template. A template is any nonzero set of medial axis elements. Our Medial Axis Model, described ahead, uses templates to compare with other stimuli (specifically, their medial axis components). In this methodology, all medial axis elements that comprise a template come from one single parent stimulus. But the template itself stands apart conceptually from the parent stimulus (i.e. the template can be compared with the parent stimulus it came from. All the compared medial axis components will be a "perfect fit" because they are identical. But conceptually there is no difference between comparing template and parent stimulus versus comparing template and any other stimuli). Revisiting the first graph of Figure 3.2b, displayed is one example template overlaid over an example stimulus. The medial axis elements that comprise the template are all drawn from the stimulus in the graph, and furthermore this particular template is the exhaustive set of all the medial axis components of the parent stimulus. The second graph in Figure 3.2b shows another template defined from the same parent stimulus. This time, the medial axis components are not exhaustive, but rather a subset of the set of medial axis components that comprise the parent stimulus.

3.2.2.3 Comparison of Medial Axis Elements

The templates described in the previous section are defined for the purpose of comparison to other stimuli to predict neural firing rates. Following is a description of our mathematical model used to make comparisons. First, a similarity index is used to compare any two single medial axis components. The index results from using a multi-dimensional gaussian kernel, varying over several dimensions. The dimensions are closely related to the generation of stimuli.

To compare medial axis elements in our similarity index, we use the direct parameters or values in the stimulus generation procedure (see section 2.2.2 Stimulus Generation). For example, location values of the skeleton precursors to limbs, junctions and terminators are considered equal to the location values of the final corresponding medial axis elements. Only medial axis elements of the same type are compared (i.e. terminators-to-terminators but not terminators-to-limbs are compared). The dimensions of the kernel vary across different medial axis type and are summarized in Table 3.1.

For terminators, there are four dimensions: two location dimensions (x and y), one of width (taken directly from the generation procedure), and one of orientation. Width is taken directly from the node width from the generation procedure. Two different types of locations were considered. One was absolute location, which was taken directly from node locations from the generation procedure. Another location type considered was relative location in which the absolute locations (x and y) were

normalized to a range of 0 to 1 corresponding to its relative position within a conceptual box containing the stimulus. This absolute versus relative location type applied to all elements (terminators, junctions, and limbs). All templates were defined as either absolute or relative. That is, all elements in a template had their locations defined the same way. Orientation of a terminator was defined using the appropriate skeleton-limb orientation from the generation procedure. However, while skeleton-limb orientation is of range [-90, 90), terminator orientation was adapted to be in range [-180,180). Comparisons between two terminators involve taking the difference between values of the corresponding dimensions. Since all these values are scalar, this is a straightforward step.

Junctions have the same dimensions as terminators, but the "angle" dimension is treated differently with respect to other medial axis types. "Angle" for a particular junction refers to the angles of the limbs radiating from that element. Since junctions, by definition, have multiple limbs radiating from them, this value is non-scalar. Defining the "order" of a junction to be the number of limbs radiating from them, the "angle" of a two-order-junction has two values, while a four-order-junction would have four values in "angle" (The maximum order of junctions was set at four). Two junctions to be compared may be of different orders. A combinatorial search for the best possible set of one-to-one orientation matches was executed. For example, in a comparison between two three-order junctions with angles at [0, 45, 180] and [50, 5, 200] respectively, angle comparison pairs of [0, 5], [45, 50], and [180, 200] would be

formed. In a comparison between a two-order and four order junction with angles [0, 90] and [45, 100, 200, 350], angle comparison pairs of [0, 350] and [90, 100] would be formed. Once the comparison pairs are formed, we take the difference between them to form a difference vector (in the examples above, difference vectors of [5, 5, 20] and [-10 10] respectively would be formed). We then characterize the difference vectors by its mean and sample standard deviation (In the above examples, the [mean, std] parameters would be [10, 8.66] and [0 14.1] respectively). The mean of the difference vector indicates a "global orientation difference" and can be interpreted as "The minimum global rotation angle that could be applied to one junction – without changing individual limb angles within the junction – that results in the best match to another junction, where 'best match' means the smallest mean squared error of the resulting difference vector". The sample standard deviation of the difference vector indicates the aforementioned mean squared error after a global rotation is applied. Since Junction orientation comparisons supply two comparison parameters (mean and standard deviation), Junction comparisons ultimately have five dimensions (whereas terminators have four).

For limbs there are five dimensions: two location dimensions (defined as the average of the locations of the two corresponding endpoint nodes from the generation procedure), one of width (also the average of the node values), one of angle (taken directly from the corresponding skeleton-limb in the generation procedure), and one of length. Unlike terminators and junctions in which angle is of range [0 360], limb

angles are of range [0 180]. All the values are scalar and so comparison of two limbs are straightforward.

Table 3.1: The Considered Dimensions of Each Medial Axis Element

	Dimensions	**Notes**
Terminators	• Location (x) • Location (y) • Width • Orientation	• All locations (all element types) could be absolute or relative. Each Medial Axis Model consisted of elements defined entirely absolute or entirely relative. • Term and Junc Orientation range: [-180, 180]
Junctions	• Location (x) • Location (y) • Width • Orientation (Vector)	• Two parameters capture the "difference" between two Orientation vectors: "Global Orientation" and "Individual Deviation" difference. Two junctions of different orders may be compared.
Limbs	• Location (x) • Location (y) • Width • Orientation • Length	• Width and Location values are calculated as the average of the corresponding values of the two endpoint nodes. • Limb Orientation range: [-90, 90]

3.2.2.4 Predicted Firing Rate: Model Capture and Sigmoid Function

Each pair of medial axis elements to be compared are passed through a gaussian kernel with the comparison dimensions defined above. The resulting values are "individual capture" values which range from 0 to 1 (with 1 signaling a pair of medial axis elements that are identical).

A set of individual capture values (resulting from comparisons of multiple pairs of elements) are averaged to obtain the "general capture" value (simply referred to as "capture" from now on), which again ranges from 0 to 1.

$$capture = \sum_{i=1}^{n_m} \exp(-(\sum_{d=1}^{n_d} \frac{(m_{i,d} - t_{i,d})^2}{2\sigma_d^2}))$$

where each i iterates through the pairs of medial axis elements (one from the template, one from the stimulus) to be compared, and d iterates through the dimensions of comparison of each pair.

Capture values are generated by comparing a template to a comparator stimulus. The template and stimulus together provide two sets of medial-axis elements to be compared. One-to-one pairs of individual medial axis elements (one from template, one from stimulus) have to be formed, but it is unknown a priori the optimal mapping from one set to the other. Therefore, a combinatorial sweep of all possible pairing of elements (of matching element type). For example, if the template set had 3 limbs and the comparator set had 2 limbs, then there were 3-choose-2 possible sets of limb comparisons that could be made. The same goes for terminators and junctions and the total number of element mappings possible would be (# of terminator pairings) * (# of junction pairings) * (# of limb pairings). Capture values of all possible mappings were ascertained, the maximum was ascertained, and the mapping that resulted in that maximum capture value was saved.

To recap, a Medial Axis Model defined by a single template composed of medial axis elements, which can belong to one of three types: Terminators, Junctions, and Limbs. The template can then be compared to any stimulus, generating a capture value associated with that stimulus. As a final step, The Medial Axis Model then

transforms the capture values via a sigmoid function to generate the final predicted firing rate of that stimulus.

Figure 3.2c visualizes example similarity captures. The exhaustive template from Figure 3.2b is compared to two stimuli (graphs on the left) and the resulting element-by-element captures are visualized by the graphs on the right with the green scale displaying increasing levels of element capture. Each element highlight corresponds to one gaussian kernel value which ranges from 0 to 1 (colored from black to green respectively).

3.2.3 Constraining the Medial Axis Model

3.2.3.1 Free Parameters and Model Initialization

The free parameters of this model are the sigma terms of the gaussian kernels (one for each of the comparison dimensions) and those of the sigmoid function (four parameters: upper and lower plateau, inflection point, and first derivative at the inflection point). To generate models, a set of "baseline" sigmas are established. Each baseline sigma was set at roughly one quarter of the range of the corresponding dimension. For example, the baseline sigma of the terminator orientation dimension was set at 90 degrees which is one quarter of the 360 range of orientation. This allows any template to generate a set of (initial) capture values upon comparison with all the recorded stimuli (of that particular cell). To generate a set of (initial) predicted

firing rates, the free parameters of the sigmoid function are determined by using least squares regression to fit the set of initial capture values to the set of recorded firing rates.

3.2.3.2 Template Search Methodology

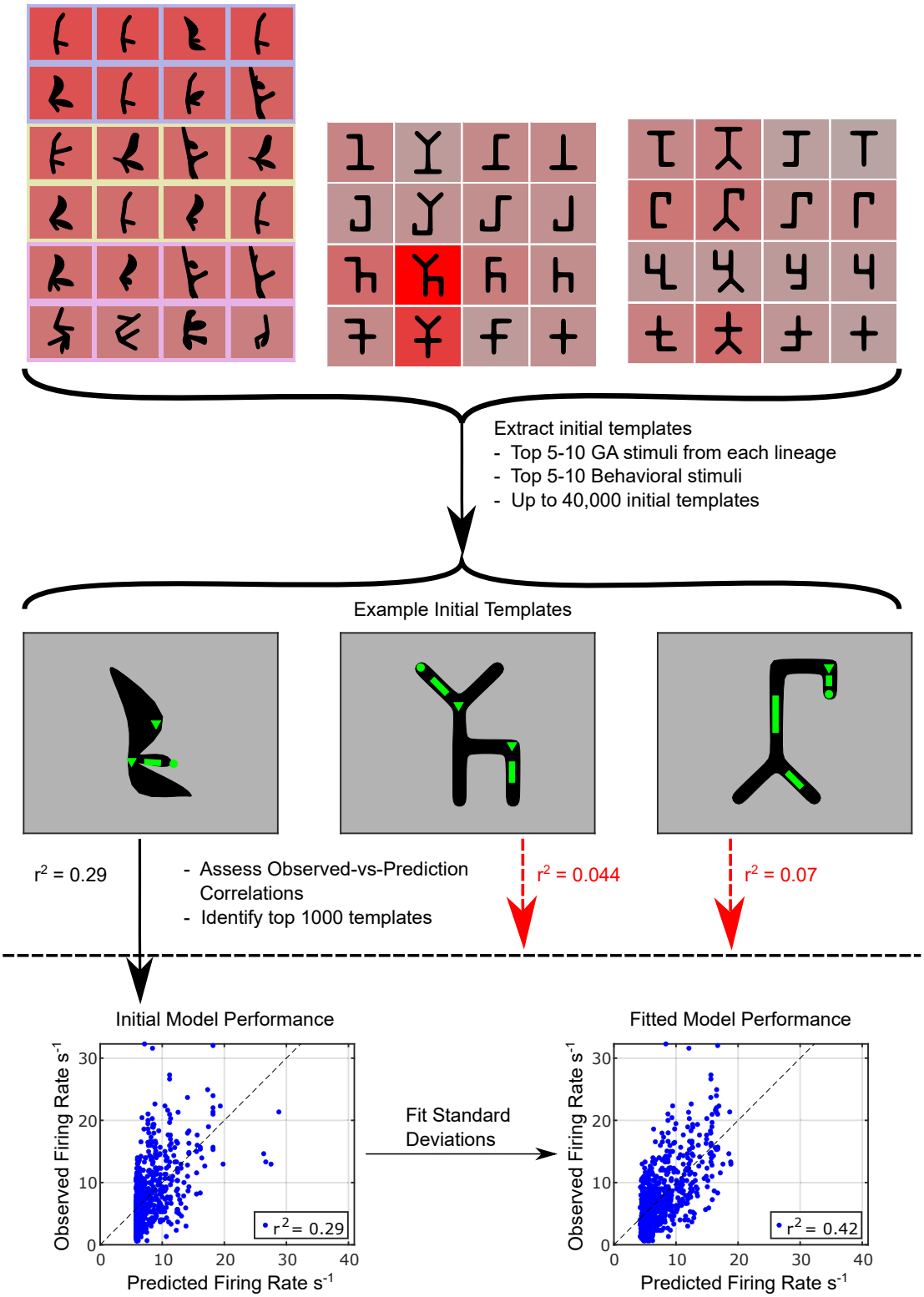
To find the best template that best explains neural behavior, we perform a search for candidate templates from a pool of high-response stimuli (Figure 3.3). Specifically, 5 to 10 stimuli from each GA lineage and 5 to 10 of the top Behavioral stimuli were selected to seed templates. From this pool, an exhaustive set of all possible templates were extracted. Templates ranged from single-element to the entire parent stimulus from which they came. The resulting set of templates number up to roughly 40,000. From this initial set of templates, we calculated the predicted firing rate as described above. Correlations between predicted and actual firing rates were calculated, and the top 1000 templates were selected for further analysis. A least squares fit was executed to optimize the standard deviations of the similarity equation. The sigmoidal function was also re-fitted and new predicted firing rates and correlations were assessed. The top 10 templates of each of the seven element types (see section 3.2.4.2.2 Element Importance) were then saved, for 70 total final templates.

Figure 3.3: Aim 2 Methods: Template Search

5 to 10 stimuli from each GA lineage and 5 to 10 of the top Behavioral stimuli were selected to provide candidate templates. From this pool, an exhaustive set of all possible templates, ranging from single elements to entire parent stimuli, were extracted. The number of initial templates could number up to 40,000. From this initial set of templates, we calculated the predicted firing rate across all shown stimuli using baseline standard deviation values (similarity index) and sigmoid function. Correlations between predicted and actual firing rates were assessed and the top 1000 templates were selected for further analysis. In the figure, an example of 3 initial templates are displayed, with only the first template exhibiting a high enough correlation to warrant further study (dotted black line). A scatterplot of the initial observed versus predicted firing rates is shown on the bottom left. All selected templates are then optimized. The free parameters (standard deviations and sigmoid parameters) are fitted by least squares regression to obtain the final model. The scatterplot of the fitted model is shown on the bottom right. The top 10 templates of each of the seven element types (see Element Importance) were then saved, for 70 total final templates.

(next page)

Figure 3.3: Aim 2 Methods: Template Search



3.2.4 Exploring Learned Effects

3.2.4.1 Learning Threshold

Because we did not record the full set of protocols for all the cells (see Table 2.1), we could not directly observe a learning signal from all the cells. The direct learning signal would have been to compare 2-way F values (specifically bottom and interaction) between Canonical and Inverted stimuli. But nevertheless, we tried to use the limited information we had to make a best guess at the learned signal of the incomplete cells.

The signal we have from all the cells is one-way F values across the 8 Trained stimuli. Figure 3.4a displays one-way Trained F values versus the sum of two way Canonical F values of Bottom and Interaction (henceforth referred to as "$F_{One-Way}$" and "F_{BIsum}" respectively). Blue data points represent full-protocol cells, while red data points are non-full-protocol cells in which F_{BIsum} was not observed. First we note that amongst the observed blue data points, $F_{One-Way}$ is significantly correlated with the canonical F_{BIsum} ($r^2 = 0.51$, p = 5.9e-10). By comparison, the $F_{One-Way}$ versus inverted F_{BIsum} (not shown) are almost completely independent.

We use this as justification for using $F_{One-Way}$ as an estimator of canonical F_{BIsum}. The red data points in Figure 3.4b show data for non-full-protocol cells. The $F_{One-Way}$ values are as recorded, but the canonical F_{BIsum} values are predicted based on the observed relationship across full-protocol cells. Furthermore, we can define two thresh-

old lines for the two F variables respectively which breaks the Figure 3.4a graph into quadrants. Of note is that the lower right quadrant is empty. This means that none of the recorded full-protocol cells with $F_{One-Way}$ values higher than 7.5 had any F_{BIsum} values less than 3.5. The significance of this relationship is apparent when considering Figure 3.4b. Here, inverted versus canonical F_{BIsum} is plotted, which is the direct learning signal that we are interested in. Also plotted are the equality line and the same threshold of 3.5 for the canonical values. Of note, for the 43 cells in which canonical F_{BIsum} is greater than the 7.5 threshold, all but two cells exhibited the learning signal of canonical F_{BIsum} ¿ inverted F_{BIsum}. Thus, we conclude that 1) cells with a $F_{One-Way}$ greater than 7.5, canonical F_{BIsum} is greater than 3.5; and in turn, 2) this furthermore implies that canonical F_{BIsum} ¿ inverted F_{BIsum}, which is the learning signal. Therefore, we can use the $F_{One-Way}$ threshold of 7.5 as a proxy threshold to separate learned versus unlearned cells. Using this threshold, we separated our population of recorded cells into 65 learned and 63 unlearned cells.

3.2.4.2 Metrics For Testing Learned Effect

We used the Learning Threshold described above to categorize cells as Learned or Unlearned. We then tested for a learned effect across the groups by ascertaining three metrics: Model Performance, Element Importance, and Spatial Contribution.

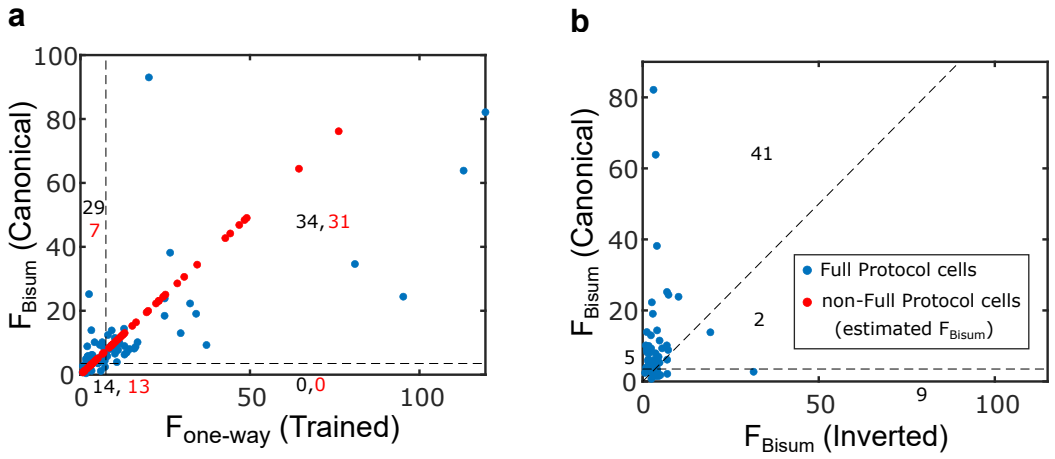


Figure 3.4: Here we establish a "learning threshold" by which we can characterize all recorded cells as "learned" or "unlearned". This includes cells that were not Full Protocol cells, meaning that comparisons between Canonical and Inverted responses could not be made directly. In (a), we compare Trained $F_{One-Way}$ (see Aim 1-part 2), versus Canonical F_{BIsum} (see Aim 1-Part 1). Full-Protocol cells are represented in blue data points, and there exists a significant correlation between the two variables. This is used as justification to be able to estimate the remaining two-way ANOVA data among the non-Full Protocol cells via linear regression. The estimated data is represented by the red data points. The dotted lines represent two thresholds, one for each variable. The thresholds break up the space into quadrants. Numbers corresponding to each quadrant indicate how many cells (Full Protocol or non-Full Protocol) are fall in each quadrant. Of note is that the lower right quadrant is empty, meaning that none of the recorded Full-Protocol cells with Trained $F_{One-Way}$ values higher than 7.5 had any Canonical F_{BIsum} less than 3.5. The significance of this relationship is apparent in (b), where Canonical versus Inverted F_{BIsum} is plotted. Only Full Protocol cells are plotted here. The two dotted lines are the equality line, and the same 3.5 threshold from (a). Of note, for the 43 cells above the 3.5 threshold line, all but two cells are above the equality line. Thus, we conclude that cells displaying high Trained $F_{One-Way}$ values are very likely to have Canonical F_{BIsum} values over 3.5, which in turn strongly indicate greater Canonical F_{BIsum} value over Inverted F_{BIsum} value. Therefore, we define the Trained $F_{One-Way}$ threshold of 7.5 as the "learning threshold". We categorize the 65 cells that exhibited values higher than 7.5 to be *Learned cells*. The other 63 cells are considered *Unlearned cells*.

3.2.4.2.1 Model Performance

Model Performance is straightforward. It is simply the correlation of observed-versus-predicted firing rates of genetic algorithm stimuli. The predicted firing rates come from the best Medial Axis Model of that cell. Typically this best model had a template that consisted of all element types (see next section: Element Importance).

3.2.4.2.2 ElementImportance

For each cell, we ascertained the relative importance of each element type (terminators, junctions, and limbs) with respect to model performance. A high importance of an element type meant that inclusion of that element type was crucial to model success. All medial-axis models belonged to one of seven types: 1) All element types, 2) No Terminators, 3) No Junctions, 4) No Limbs, 5) Only Terminators, 6) Only Junctions, or 7) Only Limbs. From these seven types, model performance associated with each type is ascertained by taking the mean correlation (r, not r^2) across the 10 top performing models of that type. In the Template Search Methodology, it is stated that 70 final medial axis template models are saved: 10 each of seven element types. Finally, element importance is ascertained by taking the difference between all-element performance (that is, mean correlation across the 10 top performers) and performance of the appropriate other type of model. For example, Terminator importance is ascertained by subtracting No-Terminator performance from All-Element performance. We could also ascertain double-element importance. For example, Junc-

tion+Limb Importance is ascertained by subtracting Only-Terminator performance from All-Element performance. It should be noted that All-Element Performance was consistently greater than all other element-type performance, and so Element Importance values across cells and type were always positive (with very rare exceptions).

Figure 3.5a shows some example model types of one cell. For illustrative purposes, only types 1-4 are shown. Each data point in the four graphs compares the observed firing rate of a GA stimulus, with the average predicted firing rate of that element type. That is, the average predicted firing rate across the 10 top templates of the element type in consideration. The resulting correlation (performance) between observation and prediction is shown. Also each graph has an inset, which shows an example of one (out of 10) of the models of that particular element type.

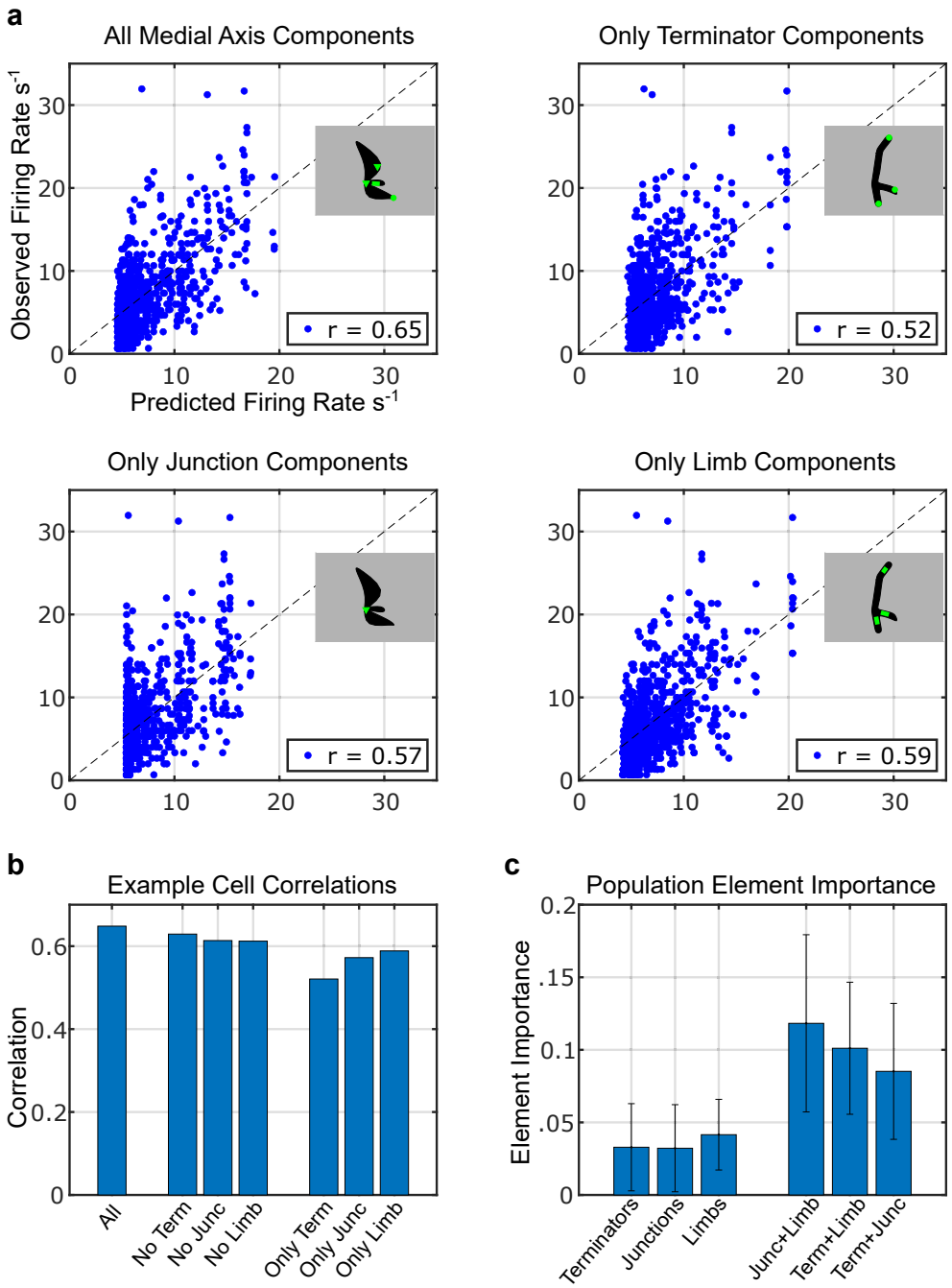
Figure 3.5b shows the model performance of all seven element types for this example cell. Note that All-Element performance is the highest, while two-element models outperformed one-Element models. From these performance values, six Element Importance values for this example cell can be obtained by subtracting out of All-Element-performance the rest of the six performances. Figure 3.5c shows the resulting Element Importance values when this is executed across the population of cells. Single-Element importance (terminators, junctions, and limbs) are, not surprisingly, smaller than Double-Element Importance (Junc+Limb, Term+Limb, Term+Junc).

Figure 3.5: Aim 2 Methods: Element Importance

Towards constructing the Element Importance metric: the importance of each medial axis element type to the performance of Medial Axis Models is ascertained across cells. a) Four out of seven (see Methods) model element types are shown: All Elements, Only Terminators, Only Junctions, and Only Limbs. The data points in each graph show predicted versus actual firing rate, where the predicted firing rate is the average of the top 10 performing models of that element type. In the insets of each graph, the template of the top performing model is displayed (only one out of the 10 templates are displayed). The metric made use of correlation (r, not r^2) and so that is displayed as well. b) The correlations across all model types for this example cell are displayed. Note that All-Element performance is the highest. Also, two-element models outperformed one-element models. The final Element Importance metrics are obtained by subtracting out the six non-All Element performance from the All Element performance. c) The resulting Element Importance values across the population of cells. Single-Element importance values (obtained by subtracting All Element performance with double-element performance) are smaller than Double-Element Importance values.

(next page)

Figure 3.5: Aim 2 Methods: Element Importance



3.2.4.2.3 Spatial Contribution

We sought to break down Model Performance into separate spatial components. Since our experimental design involves top and bottom parts comprising Behavioral stimuli, we ask whether the learned model template captures of stimuli exhibit any sort of top versus bottom trendi. To do this, we focus on the model performance over Behavioral responses, referred to as Behavioral Model Performance. We ascertain Spatial Contribution in a similar way to Element Importance. We start with the original Behavioral Model Performance and subtract out "dropped" Behavioral Model Performances where we drop contributions from either tops or bottoms.

Specifically, we start with the top 10 templates (utilizing all element types) that comprise the original Behavioral Model Performance. As described in the Medial Axis Template Section, any given template generates a predicted firing rate for any particular stimulus by 1) generating a set of individual capture values (this involves a subset of the medial axis elements of the stimulus), 2) averaging the set of individual capture values to obtain the general capture value, and finally 3) transforming the result through a sigmoid function. In Figure 3.6a, capture sets from some Behavioral Stimuli are shown. Note that the individual captured elements span top and bottom elements of each of the stimuli in this example.

In order to ascertain the importance of a spatial side (i.e. top elements) to the Behavioral Performance, first note that every element in a Behavioral stimulus can be categorized as a "top" or "bottom" element because each stimulus by definition is

fully described by one top and one bottom part. For this analysis, we introduce an extra "drop" step after the individual capture step (1). We remove from consideration individual capture values from which the corresponding elements belong to one of the two spatial sides (top or bottom). Figure 3.6b shows examples of dropped bottom and top captures respectively. Each panel in Figure 3.6b are dropped versions of the corresponding panels in Figure 3.6a. Figure 3.6c is the same, but for Bottom-only capture.

After dropping one spatial side, the new sets of capture values are summed and inputted through a sigmoid function (as they were before for the original model procedure) to generate predicted firing rates. The parameters of the sigmoid function are refitted by least-squares regression to correctly map the range of the new capture values onto the range of observed firing rate (and refit the nonlinearities of a sigmoid function). The scatterplot in Figure 3.6d shows the observed versus predicted firing rates for original, top-only, and bottom-only Behavioral Models. Note that for this particular cell, top-only predicted firing rates roughly mimics original-capture predicted firing rates. However bottom-only capture fails to predict firing rate with any accuracy. The observed-vs-predicted correlations shown in the graph reflects this effect. Two metrics result: $Spatial_{Top}$ and $Spatial_{Bot}$. Each metric is defined by the difference between the original observed-vs-predicted correlation and the corresponding dropped-side correlation. The bar graph in Figure 3.6d shows the population distribution of these metrics, with the red lines highlighting the example in the above

panels. Collectively these two metrics are referred to as General Spatial Contribution. General Spatial Contribution in the end was not very informative.

Extending this analysis further, we also calculated individual Spatial Contribution metrics for Canonical and Inverted stimuli. Instead of starting with the observed-vs-predicted correlation across all stimuli, we would only consider the correlation over Canonical or Inverted stimuli. Dropped capture values and subtraction of dropped-side correlation was executed the same as before. What results are four metrics: $Spatial_{Top,Can}$, $Spatial_{Bot,Can}$, $Spatial_{Top,Inv}$, $Spatial_{Bot,Inv}$. Figure 3.6e shows the population distribution of these metrics, with the red lines highlighted the example in the above panels. Collectively these metrics constitute the final "Spatial Contribution". These metrics in combination with the Learning Threshold ultimately extend our results from Aim 1.

Figure 3.6: Aim 2 Methods: Spatial Contribution

Towards constructing the Spatial Contribution metric: a measure of the importance of top versus bottom captured elements towards the overall model performance. a) Displayed are a subset of Canonical and Inverted stimuli, each with highlighted individual captured elements following comparison to an example template. (b,c) Displayed is the same template capturing the same subset of stimuli, but with one spatial side (top or bottom) "dropped". Individual captures of elements belonging to either the top (c) or bottom (b) side are set to zero. What remains are the individual captures belonging to the other side. Note that for this template, more top elements seem to be captured instead of bottom elements. It would seem reasonable to predict, then, that this template carries more information in top elements. To quantify this, predicted firing rates for all three contexts (both sides, top-only, and bottom-only) are obtained. The predicted firing rate for both-sides is simply the original model's firing rate. The predicted firing rates for top- and bottom-only are obtained by refitting the sigmoid function parameters to match the new set of capture values with the observed firing rate. The results are plotted in the scatterplot in (d), along with their correlations. As expected, the bottom-only context performs the worst out of the three. Top-only and bottom-only correlations are then subtracted from original (both-sides) correlation to obtain two metrics: $Spatial_{Top}$ and $Spatial_{Bot}$. The bar graph displays the population (all cells) of $Spatial_{Top}$ and $Spatial_{Bot}$ metrics, with the red dotted lines indicating the individual values for the example cell. These two metrics collectively constitute General Spatial Contribution. However, this would not prove very informative in the end. For (e), we repeat the procedure, but with separate and individual consideration for Canonical stimuli only and Inverted stimuli only. The scatterplot in (e) plots data only over the 16 Canonical stimuli and not all 32 stimuli. The scatterplots in (d) and (e) may look similar, but close inspection shows "extra" data points in (d) versus (e). New contribution metrics are calculated the same way as described for (d). A scatterplot for Inverted Stimuli Only is not shown. The bar graph in (e) shows the four metrics: $Spatial_{Top,Can}$, $Spatial_{Bot,Can}$, $Spatial_{Top,Inv}$, and $Spatial_{Bot,Inv}$, which collectively constitute the final "Spatial Contribution". These metrics in combination with the Learning Threshold ultimately extend our results from Aim 1.

(next page)

Figure 3.6: Aim 2 Methods: Spatial Contribution

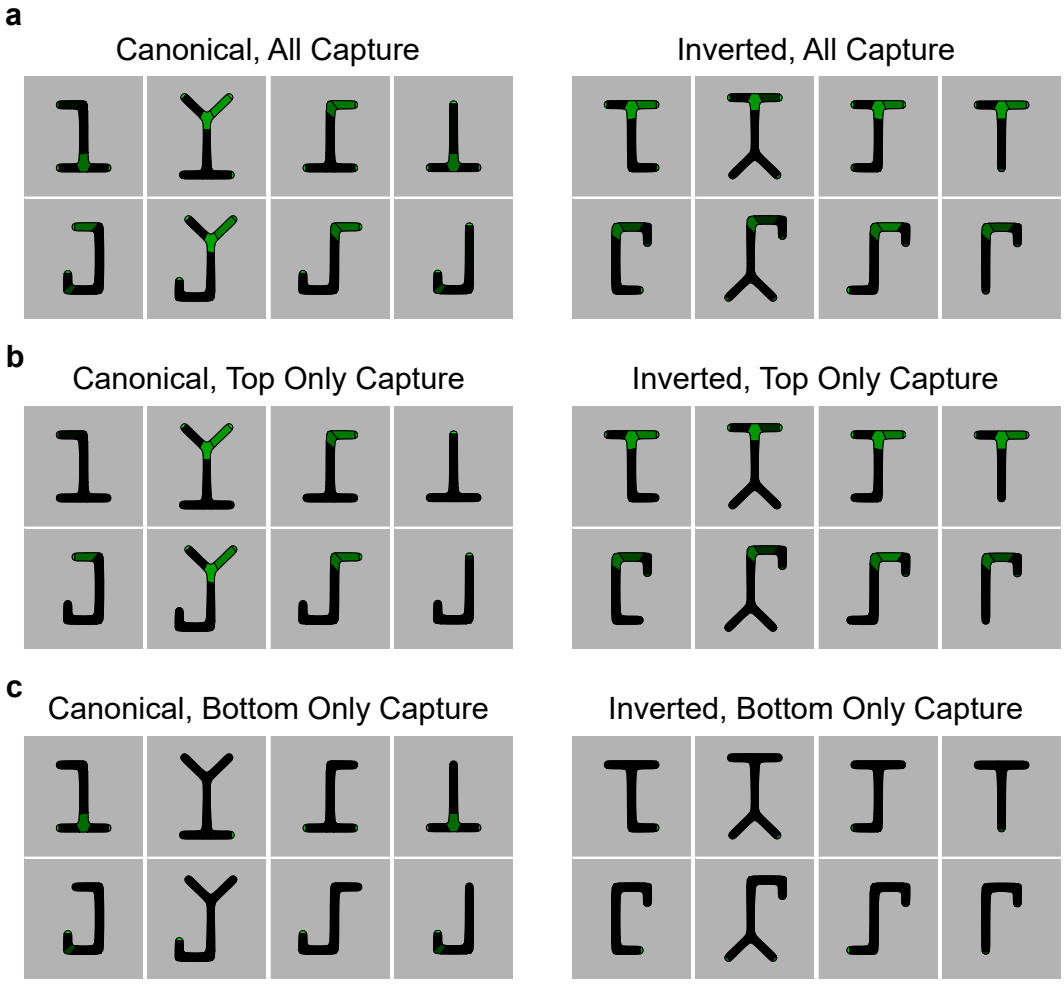
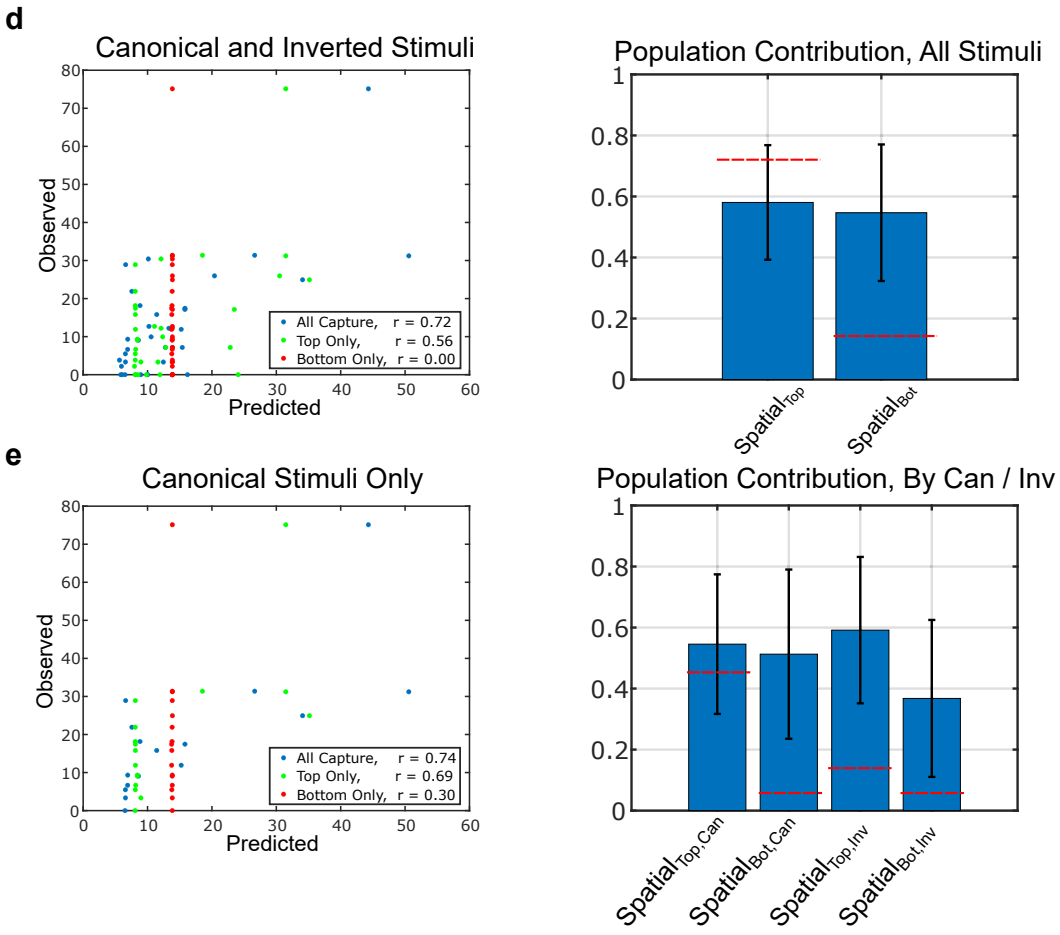


Figure 3.6: Aim 2 Methods: Spatial Contribution (Cont)



3.3 Results

3.3.1 Genetic Algorithm

Table 3.2 displays genetic algorithm (GA) duration data that was executed over 128 recorded cells. At least 5 generations were executed from 107 cells, which was the minimum criteria for future study. In general, 10 generations was the goal, and 57 cells were successfully recorded with at least 10 generations. Over all cells, a median of 9 and maximum of 23 generations were recorded.

Figure 3.7 shows how the GA successfully explores the shape space of each cell. Compared is the first and last generations of each cell's GA. Both the maximum firing rate and the range of evoked responses increase dramatically from the start to the end of the GA.

Figure 3.8 show stimulus responses from two example cells. Both GA to Behavioral responses are shown. The first cell was highly selective for the "upside-down T" shape, and the GA recovered this shape well. The second cell was selective for inverted shapes (though less selective than the first example cell was amongst Canonical shapes) and once again, the GA recovers the "upside-down h" shapes well. Once again, it should be noted that the GA was completely agnostic to the behavioral shapes. They converged on behavior-like shapes completely on their own.

Table 3.2: A Breakdown of Successfully Recorded Genetic Algorithm Generations by Cell

	Both Monkeys	**Monkey 1**	**Monkey 2**
All Recorded Cells	128	87	41
$<$ 5 Generations	21	14	7
$\geq$ 5 Generations	107	73	34
$\geq$ 10 Generations	57	34	23
$\geq$ 5 Generations + Full Protocol	47	15	32

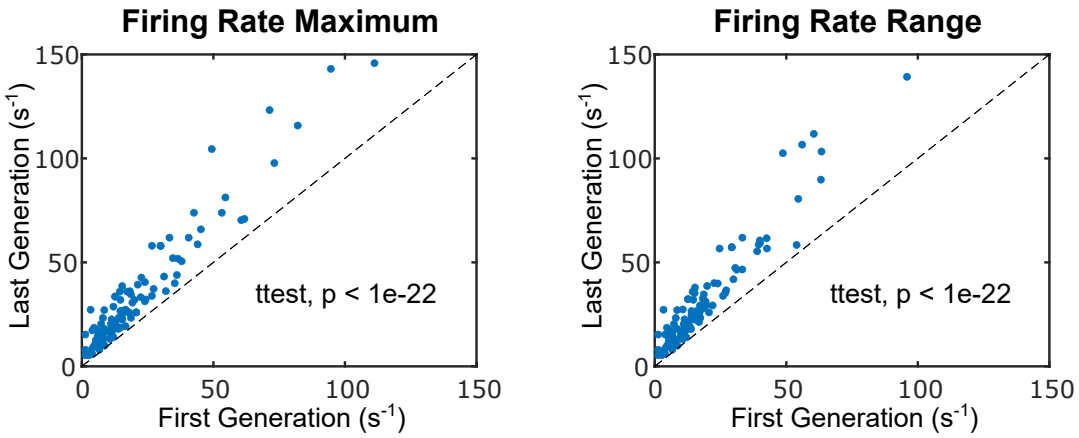


Figure 3.7: A demonstration of the genetic algorithm working as expected. Shown is a comparison of the first to the last generation data for all recorded cells. Both the maximum firing rate and the range of evoked responses clearly increase from the start to the end of the GA. Both the maximum firing rate and range of evoked responses increase dramatically (paired t-test, p $<$ 1e-22 for both variables) from the start to the end of the GA.

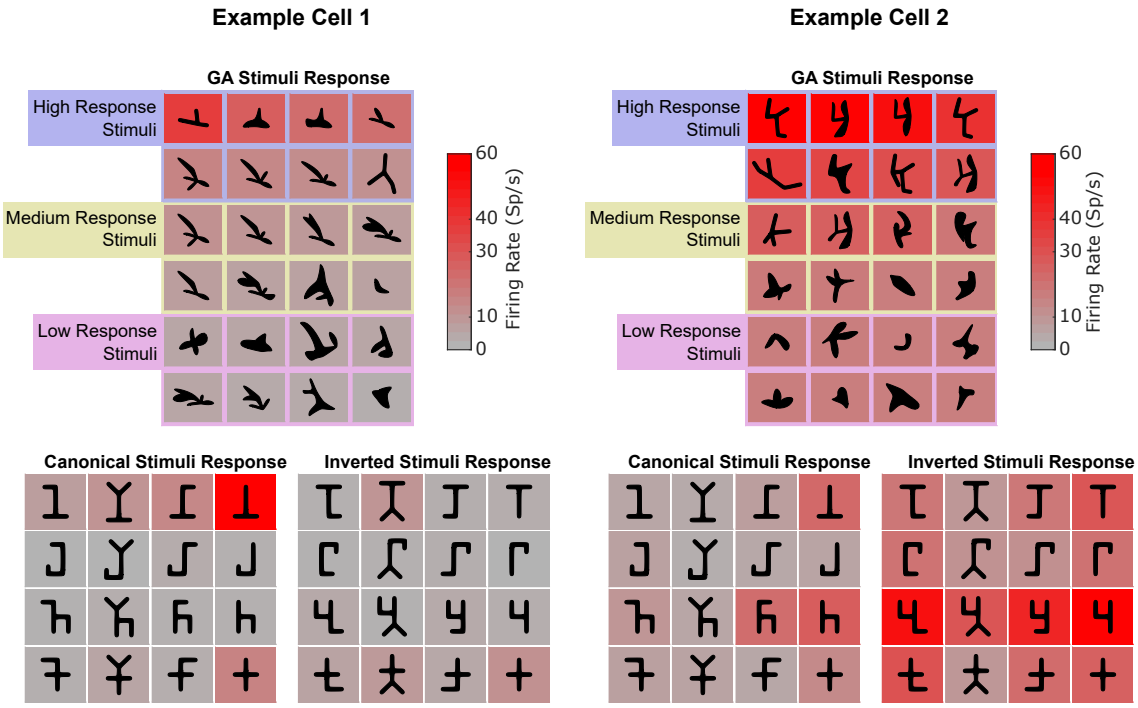


Figure 3.8: Examples of stimuli responses of two cells. Both GA (upper panels) and Behavioral (lower panels) stimuli responses are shown. One cell was selective for Canonical shapes, while the other is selective for Inverted shapes. In both cases, the genetic algorithm successfully converged on the general characteristics of both. The genetic algorithm was completely agnostic to the Behavioral shapes; convergence happens completely independently.

3.3.2 Example Model Fits

Figure 3.9 shows two examples of model fits. Shown are the medial axis template along with the parent stimulus that seeded the template, the observed-versus-predicted scatterplot with both GA and behavioral data, and the model capture of relevant behavior and GA. For each example, the Medial Axis Model performs fairly well, as shown by the observed-vs-predicted correlations.

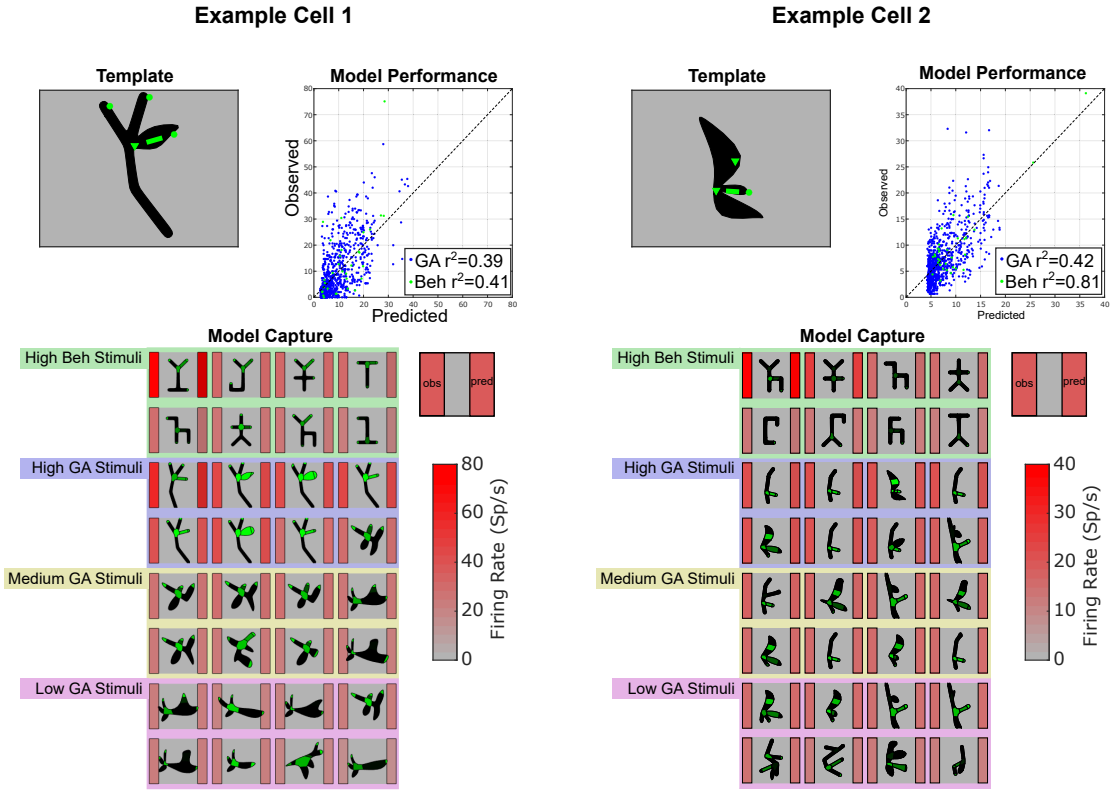
Clearly though, there is room for improvement. For example, these particular models exhibit a high amount of low-prediction/high-observed firing rates, which is not an uncommon problem among top performing models across the cells. In both examples, there are a healthy portion of stimuli that evoked high or modest firing rates, but the model predicts near zero response. This shows that there is information not captured by the model. For a discussion on potentially better versions of the Medial Axis Model, see section 4.2.1 NewGA and Template Model.

Figure 3.9: Example Model Fits

Examples of the model fits of two cells (different cells than Figure 19). Shown in the upper left panel for both cells are medial axis template along with the parent stimuli from which the template was drawn. In the upper right panels are the observed-versus-predicted comparison of stimuli responses. Both GA and Behavioral data are shown. In each example the Medial Axis Model performs fairly well, but room for improvement is absolutely present. In the lower panels, examples of behavioral and GA stimuli are shown, along with the individual capture highlights. In each stimulus panel, the left and right red backgrounds signify observed and predicted firing rates respectively.

(next page)

Figure 3.10: Example Model Fits



3.3.3 Learned Effects

We used the Learning Threshold to separate cells into "learned" and "unlearned" categories. Between these two categories, we made comparisons using three metrics: Model Performance, Element Importance, and Spatial Contribution (see Methods, 3.2.4.2). Shown in Figure 3.10 are comparisons of all the above values. Mean and standard error are shown along with sorted individual values which are color coded: Unlearned values in red and Learned in blue. Also displayed are the p-values resulting from an unpaired t-test comparing and Learned and Unlearned values.

None of the Metrics showed a clear, definitive effect. Model Performance (Figure 3.10a) had a p-value of 0.974. Amongst Element Importance (Figure 3.10b), some variables (Junction, Junction-Limb, Terminator-Junction) had a p-value less than 0.05 (0.0242, 9.33e-3, and 4.13e-3 respectively). However, visual inspection of the individual data points reveals that this effect is primarily due to a few Learned cells with large Importance values, as opposed to an actual trend across the data.

Overall, Spatial Contribution (Figure 3.10c) exhibits the best learning effect of the metrics that we test, albeit still weak. Shown in both General Spatial Contribution as well as Spatial Contribution. As with Model Performance and Element Importance, inspection of the individual data points belie weak trends rather than a striking difference between Learned and Unlearned population. Nevertheless, unlike Element Importance, the trend does not appear to be confined to only a handful of cells at the extremes; the trends seem real even if weak, so it is worth further discussion.

Among General Spatial Contribution, top contribution p-value is 0.242. The bottom contribution p-value is 1.05e-04. This fits with the Aim1 ANOVA results which showed learning effects for bottom part selectivity, but not top part selectivity. For Spatial Contribution, which divides further by Canonical and Inverted Behavioral responses (see Methods, 3.2.4.2.3), Canonical-Top, Canonical-Bottom, and Inverted-Bottom exhibited "significant" p-values of 4.03e-3, 1.44e-07, and 3.93e-2 respectively. Inverted-Top exhibited a "non-significant" p-value of 0.313. Of note, in all of the effects reported so far (Model Performance, Element Importance), the mean learned value was always greater than the unlearned value. Here however, both Inverted-Top (not-significant) and Inverted-Bottom (significant) Spatial Contribution exhibited lower learned mean.

Figure 3.10: Learned Effects Over Various Metrics

Comparisons of Learned (blue) versus Unlearned (red) cells along three metrics: Model Performance, Element Importance, and Spatial Contribution. In all panels, unpaired t-tests comparing Learned and Unlearned values are conducted, and the p-values are displayed. a) For Model Performance, no learning effect is apparent. b) for Element Importance, while some variables show a "significant" effect with $p < 0.05$, visual inspection reveals an unconvincing effect. Any effect seems to be confined to a few outlier cells. Overall, we conclude that there is no learning effect over the Element Importance metrics. c) Two versions are displayed (see Methods) separated by different background panel color: General Spatial Contribution and Spatial Contribution. General Spatial Contribution ascertains top and bottom contributions separately. Spatial Contribution makes a further separation of contributions over Canonical versus Inverted Behavioral responses. These metrics seem to exhibit learning effects, albeit weak. Unlike Element Importance, Spatial Contribution trends do not appear confined to a few outlier cells, but rather, seem to reflect a real effect.

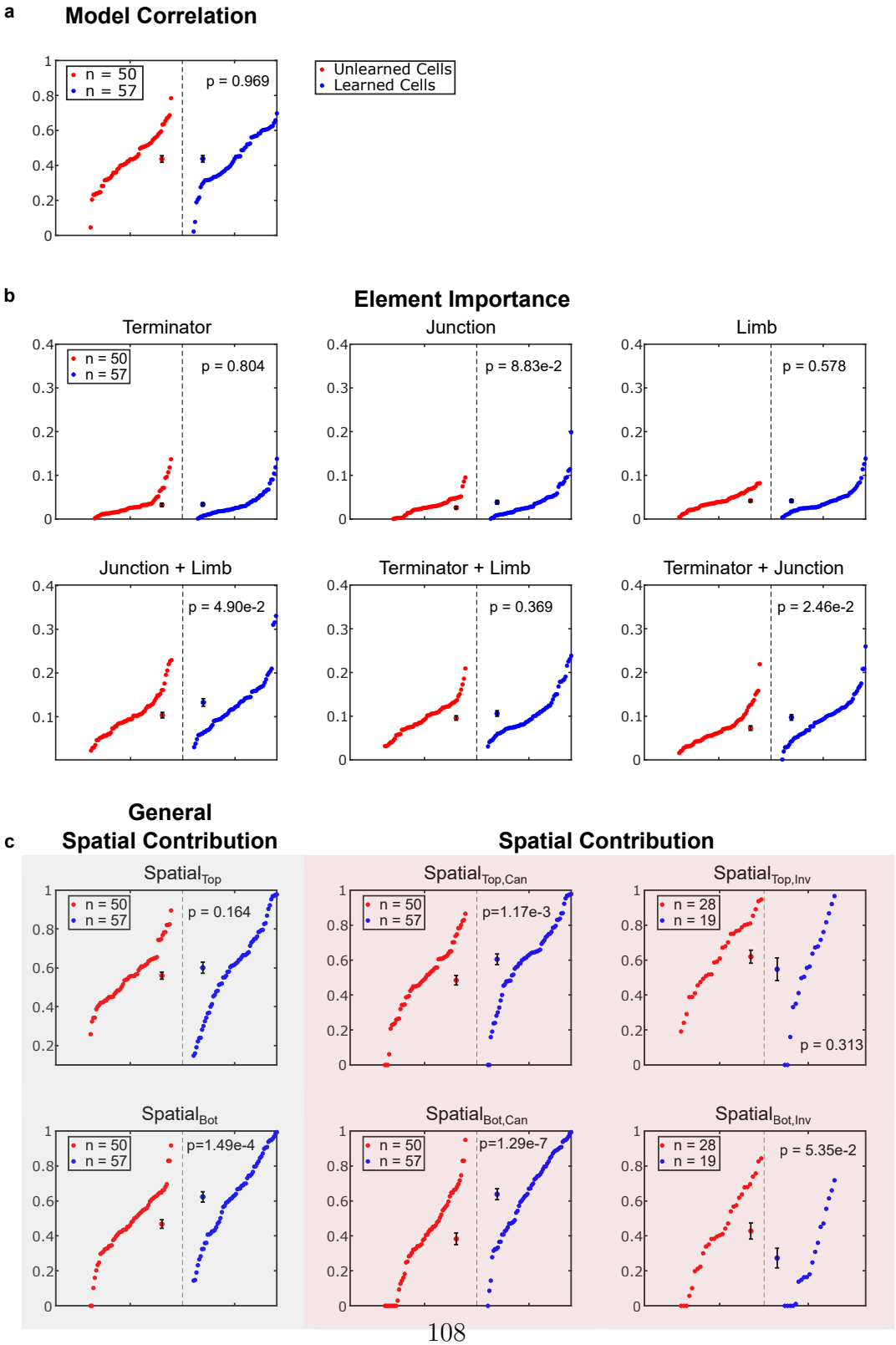
(next page)

3.3.3.1 Further Analysis of Spatial Contribution

To dive deeper into insight gained from Spatial Contribution, we first note that Spatial Contribution correlates well with the two-way ANOVA results of Aim 1. Figure 3.11 plots each Spatial Contribution metric against the corresponding (logged) main F value from the two-way ANOVA parsing part selectivity. In all cases, the correlation is significant. This demonstrates that in general, the spatial information captured by the Medial Axis Model tracks well with the spatial selectivity exhibited by the corresponding cell.

We further investigated the relationship between 8 groups of Spatial Contribution: the four Spatial Contributions metrics defined in methods section 3.2.4.2.3, further divided by Learned and Unlearned groupings. For example, one group is Learned $Spatial_{Top,Can}$. Another is Unlearned $Spatial_{Bot,Inv}$. Figure 3.12 shows eight boxes corresponding to each group. The boxes are arranged such that each group is in the same spatial arrangement as the corresponding groups in Figure 3.10c. The number in each box denotes the mean contribution of that group across appropriate cells. Comparisons between any two groups were ascertained by unpaired t-tests. This is the same as in Figure 3.10c, but expanded to all possible comparisons between groups, not just learned vs. unlearned. The boxes are color-coded in a red scale, corresponding to their relative spatial contribution. The more red, the higher the spatial contribution of that group. Notable significant differences between groups are displayed. To explain the global trends in this figure, we will discuss the trends in four

Figure 3.10: Learned Effects Over Various Metrics



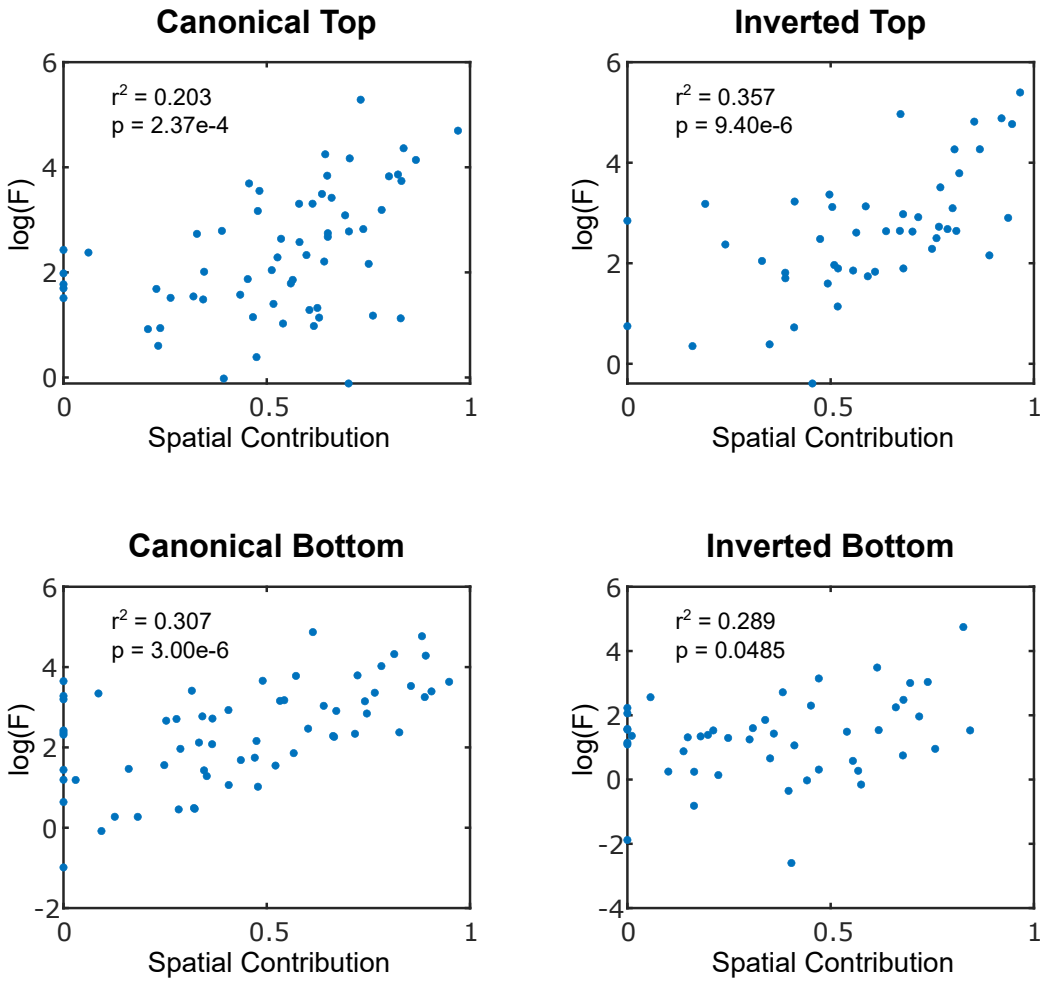


Figure 3.11: A comparison between Spatial Contribution and (logged) two-way ANOVA F values (from Aim 1). Only main effect F is used here (no interaction effects). There is good one-to-one correlation over the four corresponding levels of Canonical Top, Canonical Bottom, Inverted Top, and Inverted Bottom. This demonstrates that the spatial information captured by the Medial Axis Model tracks well with selectivity displayed by the corresponding cell across spatial region.

sub-groups individually: 1) comparisons between bottom groups only, 2) comparisons between top groups only, 3) comparisons between Unlearned groups only, and 4) comparisons between Learned groups only.

First, consider comparisons among bottom groups ($Spatial_{Bot,Can}$ and $Spatial_{Bot,Inv}$). Among Canonical: Learned $Spatial_{Bot,Can}$ was significantly greater than Unlearned $Spatial_{Bot,Can}$. Among Inverted: Learned $Spatial_{Bot,Inv}$ was less than Unlearned $Spatial_{Bot,Inv}$, though was not significant (it was close. $p < 0.10$). This result is fairly expected. Learned cells were defined such that bottom+interactive ANOVA F values were greater for Canonical over Inverted. Given that the Medial Axis Model performed fairly well and captured information from a particular side that reflects selective F values, it is reasonable to expect that Learned $Spatial_{Bot,Can}$ will be greater than Learned $Spatial_{Bot,Inv}$. Similarly, given that Learned cells were selected to have a small Inverted-Bottom F, it is a reasonable expectation that Learned $Spatial_{Bot,Inv}$ will be small, which it indeed is. Overall, there is a significant learning effect among Canonical Bottom groups, and no significant learning effect among Inverted Bottom groups. Furthermore, Learned $Spatial_{Bot,Can}$ is significantly greater than all other bottom groups.

Now consider all the top groups. The first notable result is that there is a learning effect among Canonical-Top groups: Learned $Spatial_{Top,Can}$ is significantly greater than Unlearned $Spatial_{Top,Can}$. This is interesting and not necessarily expected because the Learning threshold did not incorporate Canonical-Top ANOVA. Further-

more, the presence of learning in Canonical Tops is not a result that was evident from Aim 1 Results, in which it was concluded that learning did not happen in Tops. Inverted-Top shows no learning effect, as would be expected amongst shapes that were not trained and with a threshold that does not take Tops into account. It remains unclear why Inverted contributions are so strong relative to Canonical contributions (although this is consistent with Aim 1 results). What it amounts to is that Learned $Spatial_{Top,Can}$ is not significantly greater than all other Top groups (as would have been "ideal"), but instead, Unlearned $Spatial_{Top,Can}$ is significantly less than all other Top groups. Focusing on a specific pair: Unlearned $Spatial_{Top,Inv}$ is significantly greater than Unlearned $Spatial_{Top,Can}$ (p = 5.30e-3). This surprising finding might suggest that the inverted top medial axis elements (which are inverted versions of canonical bottom medial axis elements) inherently have more information over canonical top medial axis elements (they do, if fact, have a larger number of medial axis elements overall).

This brings us to comparison of all Unlearned groups. One possible interpretation from these comparisons is that they seem to favor Top Contribution over Bottom Contribution. This is both manifest in Canonical and Inverted responses. It is not clear why this would be the case, but taken at face value, this provides evidence that cells uninfluenced by training "pay attention" more to tops over bottoms. As mentioned above, the statistical difference between Unlearned $Spatial_{Top,Inv}$ and Unlearned $Spatial_{Top,Can}$ is difficult to interpret.

Finally, in considering the set of all Learned Contributions, the striking result is that the disparity between Top and Bottom Contribution *disappears only with Canonical responses*, and the disparity remains (grows stronger) in Inverted responses. In other words, taken collectively, the results seem to paint the picture: cells natively prefer or pay attention to top portions of stimuli, but learned cells pay attention to *both* top and bottom portions of Trained stimuli while still only paying attention to tops of Untrained stimuli.

These results both confirm and extend the results of Aim1.

Confirmation:

- Spatial Contribution traces well with two-way ANOVA main effect F values as shown in Figure 3.11.
- The learning effect manifests mainly in the Bottoms. The UnLearned $Spatial_{Bot,Can}$ to Learned $Spatial_{Bot,Can}$ difference is the largest in Figure 3.12.
- The unexpected strength of Inverted Tops versus Canonical Tops still remains.

Extensions:

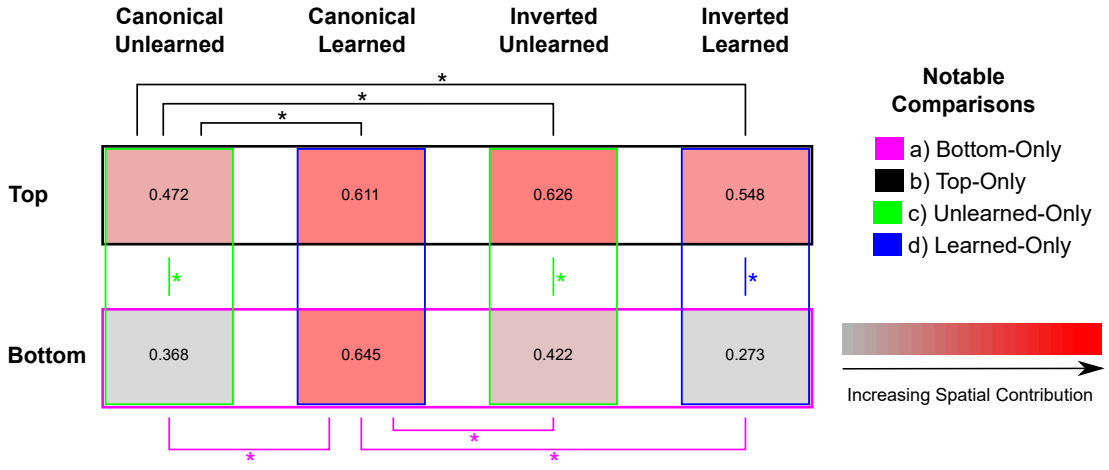
- Unlearned Cells may innately prefer tops over bottoms. This may partially explain why Inverted Tops are so large.
- Learned Cells responding to Canonical stimuli exhibit equal relative top and bottom contributions. This indicates that **learning serves to "equalize" the spatial representations of familiar stimuli**.
- Learning *does* occur for Canonical Tops, but not as strong as learning for Canonical Bottoms. Once again, this serves to equalize the spatial representations of familiar stimuli.

Figure 3.12: Spatial Contribution: A Closer Look

A graphic displaying the relationship between eight groups of Spatial Contribution: the four Spatial Contributions metrics defined in methods section 3.2.4.2.3, further divided by Learned and Unlearned groupings. The spatial arrangement of each box is the same as the corresponding groups in Figure 3.10c. The number in each box denotes the mean contribution of that group. Comparisons between any two groups were ascertained by unpaired t-tests. Significant differences are marked by the stars. Comparisons between specific subgroups are further dissected. a) The purple highlights draw attention to the four bottom groups, where Learned $Spatial_{Bot,Can}$ was significantly greater than Unlearned $Spatial_{Bot,Can}$ while Learned $Spatial_{Bot,Inv}$ was almost significantly less ($p < 0.10$) than Unlearned $Spatial_{Bot,Inv}$. Learned $Spatial_{Bot,Can}$ was the significantly greater than all other bottom groups. b) Considering top groups, we find a learning effect among Canonical Top groups. This is not necessarily expected from the results of Aim1, and we see why here. $Spatial_{Top,Inv}$ (both Learned and Unlearned) are very high, which is non-intuitive. Confusingly, Unlearned $Spatial_{Top,Inv}$ is significantly greater than Unlearned $Spatial_{Top,Can}$, suggesting perhaps, that Inverted Tops have inherently more information than Canonical Tops. However, the presence of a learning effect among $Spatial_{Top,Can}$ and not $Spatial_{Top,Inv}$ is interesting. c) Comparison within Unlearned groups of Top versus Bottom reveals a striking pattern: Tops seem to have inherently more information represented in Unlearned cells (regardless of Canonical or Inverted). d) The same Top versus Bottom comparison in Learned groups, however, shows an equality (non-significance comparison) between $Spatial_{Top,Can}$ and $Spatial_{Bot,Can}$ groups, while maintaining the Top bias when considering $Spatial_{Top,Inv}$ and $Spatial_{Bot,Inv}$ groups. This all collectively suggests that 1) Tops "start" out with more represented information in Unlearned Cells, and 2) learning boosts representation on both Tops and Bottoms, and functions to 3) "equalize" the representations of Tops and Bottoms so that in Learned cells, there is no difference between Spatial Contributions of Tops versus Bottoms.

(next page)

Figure 3.12: Spatial Contribution: A Closer Look



3.4 Previous Attempts

This section discusses previous attempts at Modeling and interpreting data. While all of these efforts were negative results (hence not the final model/interpretation), they serve to fill in the context around the current model/interpretation, which was itself mainly a negative result. In the discussion chapter, future avenues to explore will be addressed.

The previous methods discussed here are themselves a subset of all the avenues explored. They are only the most informative ones. I do not remember the entirety of the past attempts.

3.4.1 Previous Models

3.4.1.1 Surface Contours

The first model type used to fit neural data was not a medial axis model, but a surface contour model instead. Each element was a contour associated with a Terminator or Junction. Terminators each had one contour element, while Junctions had as many elements as were limbs attached to them. A gaussian kernel and sigmoid function similar to the eventual medial axis model was constructed to compare contour elements and generate predicted firing rates. The dimensions of the gaussian kernel were location, angle, and curvature. The full model was composed of multiple contour elements which were added sequentially. The first element was ascertained by initializing dozens of potential elements fitting kernel and sigmoid parameters using a least squares regression (same as in the final model), and then selecting the best element. A crucial difference between the least squares fit here and in the final model is that the parameters fitted include the means of the gaussian kernel along with the standard deviations (the final model fixed the means and did not change them). After the first element was obtained, the search for the next element commenced. The same initialization, fitting, and selection step occurred, but this time, what was fitted to was the difference between the observed and predicted firing rate of the first element. In this way, contour elements were added one at a time. Stopping conditions were a hard cap at six elements, as well as a requirement that an added element had to

improve observed-vs-predicted correlation by 25%.

This model could perform fairly well in some cases, but ultimately the scope was too narrow. Our stimuli were fairly complex, involving many contours. Also, the contours were inherently organized into more complex units. A junction, for example, had multiple contours, and if the entire junction was "important", then all or most of the contours of that junction would have to be arrived at separately to "recreate" the junction. In practice however, the model contours would be in diffuse regions of the contour-space relative to one another – they would only capture one part of a junction. The resulting mix of contour captures were often difficult to interpret as they seemed to be a hodgepodge of incomplete information from various parts of the stimulus.

3.4.1.2 Medial-Axis Model: Non-Template Version

The next major model version introduced Medial-Axis Elements. We decided to use Medial-Axis elements because it more directly described the pieces of the stimuli we were interested in. It directly represented Terminators, Junctions, and Limbs instead of potentially just pieces of them. The sequential search of elements and the fitting of gaussian kernel means were kept the same from the Surface Contour procedure described above.

While the direct representation of medial axis elements was a clear benefit from before, there were new problems. The search space was much more complex now.

Terminators, Junctions, and Limbs needed their own separate search spaces, but they were not independent from one another (a junction connected to a limb obviously had correlating properties). But mainly, this procedure still suffered from the sequential search which again struggled to arrive at interpretable complex shapes. On the one hand, the sequential search could identify separate regions of interest (for example, one terminator and one limb, separated by large distance, which are both important to describe neuron behavior) and combine them into a model, and furthermore do so with different weights (one could be positive and one could be negative). However, many cells exhibited top genetic algorithm shapes that seemed to display obvious common complex groupings of medial-axis elements such as U-shapes and S-shapes. These often involved multiple adjacent elements such as 3 limbs and two junctions. It was exceedingly rare for a sequential search to arrive at these shapes in the same way that the previous surface contour search almost never arrived at a model that reconstructed a junction. In short, the sequential search procedure seemed to arrive at local minima, and the models, while complex and had modest explanatory power, were very uninterpretable.

3.4.1.3 Final Version: Medial-Axis Model with Templates

We then arrived at the current version of the model which involves the use of templates. The main innovation of using templates is that it removes the need for a search algorithm to build a complex grouping from the bottom up. Instead, the

search algorithm is top down, with seeding from the top GA and behavioral stimuli. We "know" that the top stimuli contain the complex element groupings that can best explain the neuron's tuning, and so we endeavor to take advantage of this fact. For simplicity, we dropped both the kernel means and the weights of individual element capture from the fitted parameters of the model. Each template was therefore fixed in kernel values, and uniform in terms of capture importance across all elements. The templates also need not be among adjacent elements. This means that while the template model gains the power of being able to capture complex medial-axis element groupings (which previous attempts could not do), it still retained the potential to capture disparate parts of a stimulus (which previous models could do).

A technical hurdle of this version however was the combinatorial volume of potential templates to search through. A stimulus with n elements had 2^{n-1} potential templates. For seven elements (a moderate sized stimulus) this is 127 templates. But for a 13-element stimulus (largest possible), there were over 8000 templates. Each candidate template was then matched to every stimulus recorded for that cell, and each match itself was a combinatorial search (if a template had 2 limbs and a stimulus had 4, for example, then there were 4choose2 combinations of possible template-to-stimulus limb matches. Each element type, Terminators, Junctions, and Limbs, had their own combinatorial search). Time was a serious prohibitive factor and one that was never quite overcome. In the Discussion chapter, we will discuss future ways that this model should be extended. This includes the above technical hurdle, as well as

methodology considerations arising from the unexpected results from Aim 1.

Throughout all the versions of models, the predictive power never quite achieved a strong level. The template models achieved some modest correlations but many did not (maximum r^2 was .61, but mean and median r^2 was around 0.2). More work clearly needs to be done to truly model the behavior of neurons.

3.4.2 Previous Interpretations

3.4.2.1 Element Importance

The reader may be wondering why we fixated on Element Importance as a possible metric to further describe learning effects. There is no a priori reason necessarily to expect that learning would change the relative makeup of descriptive templates in terms of its medial axis elements.

The reason is that I mistakenly began interpretation of the data before all the data was processed. As mentioned above, the model template search was prohibitively long. When only about half of the cells were processed, I started interpreting the data. At the time, I found what seemed to be a striking relationship between Element Importance and two-way ANOVA results (part and part interaction selectivity). Specifically, it appeared that terminators were correlated with Canonical Bottom F, while limbs were correlated with Inverted Bottom F. Figure 3.13 shows a schematic of this early trend.

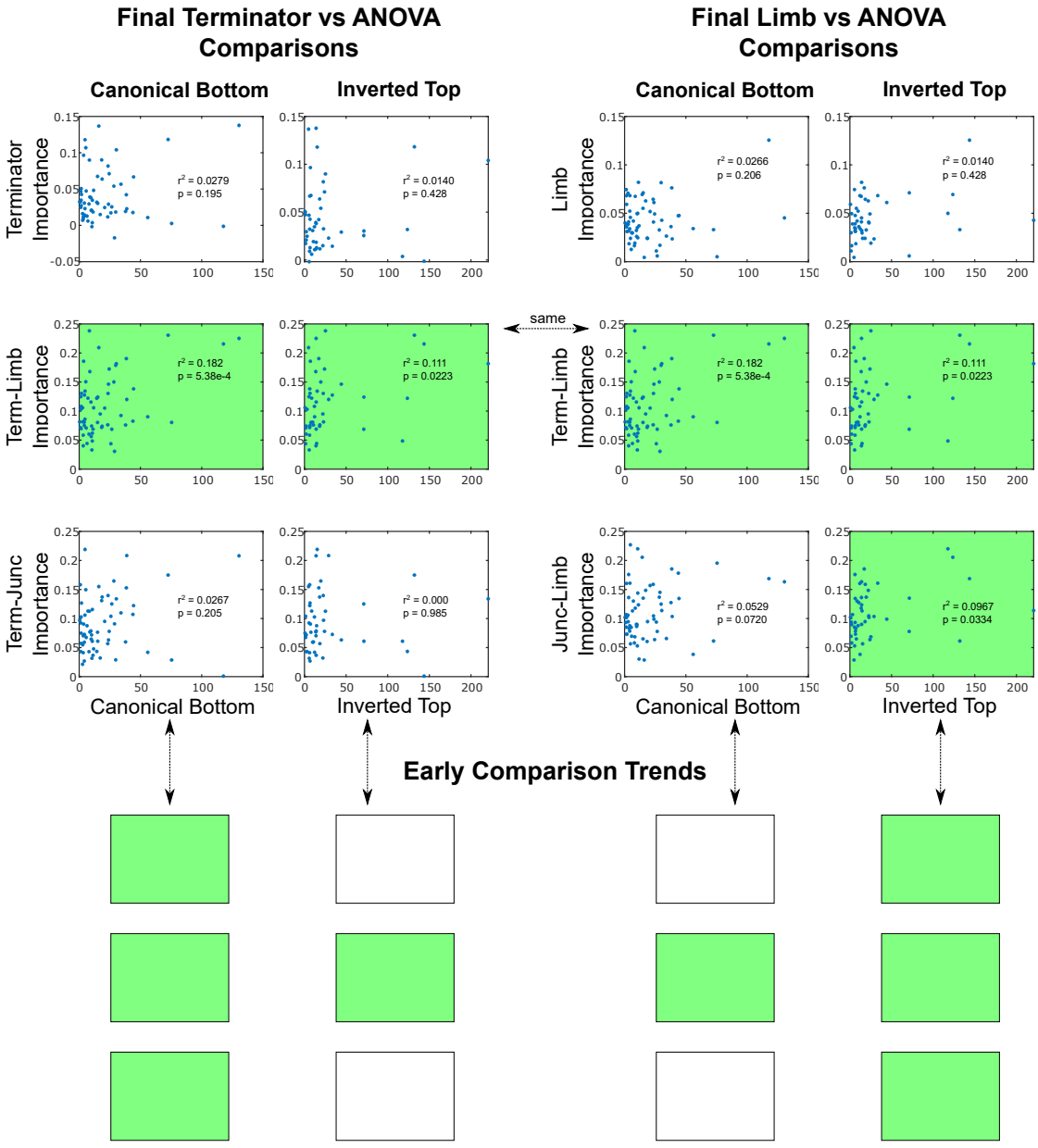
I did a number of analyses over the course of many months based on this observation, incorrectly assuming that this relationship would hold as more data was processed. However, this assumption proved erroneous as the initial relationships gradually disappeared with more cells processed. The striking separation of Terminator and Limb processing was lost. By the time we settled on the final interpretation, involving the categorization of Learned and Unlearned cells (which was moving from regression of F values to categorization of F values), all relationships between Element Importance and learning seemed non-existent or weak at best.

Figure 3.13: Element Importance versus Part and Part-Interaction Selectivity

A schematic showing comparing final and early trends in Element Importance. Each panel represents a relationship between Element Importance and ANOVA F values. This was an early test used before the Learning Threshold was developed. The left side of the graph shows Terminator companions, while the right half of the graph shows Limb comparisons. The three rows on both left and right halves show the single-element and the two double-element variables associated with Terminators and Limbs. Terminator-Limb Importance, a double element importance, is repeated on both left and right sides because it involves both Terminators and Limbs. The F values compared are both main effects: Canonical Bottom and Inverted Top Fs from two-way ANOVA in Aim 1. Displayed in each scatterplot are the correlation and associated p-value. The green backgrounds on a subset of graphs indicate significance. The bottom of the Figure shows comparison panels with the same green backgrounds indicating significance. Each panel pair from the top and bottom halves of the figure correspond to each other. Early comparison trends seemed to show a relationship between Terminator and Canonical Bottoms and a relationship between Limbs and Inverted Tops. This seemed to be very striking and led me down a rabbit hole. Ultimately though, by the time all cells were processed, this phenomena of Terminator-Limb separation was erased.

(next page)

Figure 3.13: Element Importance versus Part and Part-Interaction Selectivity



3.4.2.2 Other Descriptive Variables: Sparsity and Behavioral Threshold

All of the described efforts to correlate or categorize various metrics with ANOVA F values were repeated with sparsity (discussed in Aim 1). There was no major change in using Sparsity versus F values.

Another thing we tried was to separate the cell population based on its Behavioral Model Performance as opposed to it's GA Model Performance. The rationale was to cut out from consideration models that failed to fit the Behavioral stimulus responses. In other words, if a model failed to capture the effect that made the cell interesting in the first place, perhaps it should not be considered further. This is a different sort of threshold than the Learning threshold, which was made based on observed F values. This threshold is made based on the models performance. Thus the two thresholds could be combined to make a stricter filter: F8 Threshold + Behavioral Threshold would yield only cells that exhibited Learning Effect and whose modeling attempts successfully captured Behavioral stimulus response. Ultimately no major insights were gleaned from these efforts, and they are not included in this thesis.

3.4.2.3 Alternative Spatial Contributions

There were many attempts to describe Spatial Contribution before settling on the current method. Most of them involved direct assessment of capture values. Instead

of dropping capture values and assessing the resulting Model Performance, these early attempts simply looked at the capture values themselves. After step 1) of the capture procedure, all the capture values from one side (top or bottom) would be summed. Then various metrics were applied across stimuli. These metrics included (amongst others) averaging, standard deviations, sparsity, and kurtosis. The metrics would be applied across all stimuli in a one-way fashion or across tops and bottoms in a two-way fashion similar to two-way ANOVA. But all these methods failed the evaluation shown in Figure 3.11, which compares the Spatial Contribution metrics with the corresponding two-way ANOVA main F values. This amounts to a "sanity check", asserting that any designed spatial contribution metric should correlate with the observed selectivity of that spatial side and stimulus type (Canonical or Inverted). If a metric assigns a cell a "large top canonical spatial contribution", for example (and the template model worked correctly), it should be because that cell had large canonical top selectivity.

What made all the attempts at using direct capture values fail was the importance of the sigmoid function, which applies crucial non-linearities to the summed capture values, transforming it into the final predicted firing rate. Summing top-only or bottom-only capture values treats them linearly, but that could be wildly different from how they actually affected the model prediction. As an extreme case of this, some sigmoid functions were actually negatively sloped, meaning that the templates were actually negative templates instead of positive ones. This inverts how capture

values typically affect predicted firing rate, and metrics that do not take this into account will fail.

The final Spatial Contribution metrics used sidestep these issues by directly assessing the Model Performance instead of capture values. In this way, the nonlinear sigmoid function is incorporated into the metric as opposed to being a disconnect between the metric and the actual model prediction/performance.

Chapter 4

Discussion and Conclusion

Parametric analyses of shape coding in the ventral pathway have reliably shown that neurons throughout the ventral pathway, even in inferotemporal cortex, represent fragments of objects, or combinations of fragments [10,28,55,59,60], rather than whole object shape. However, these analyses have not generally considered whether learning could produce more holistic object representation. Baker, Behrmann & Olson generated highly suggestive evidence that extensive training could produce selectivity for whole, familiar objects. But here we extend their experimental design and ultimately challenge their interpretation.

4.1 Results Recap and Discussion

Table 4.1 summarizes all of our findings throughout both specific aims. Following is a brief recap.

Table 4.1: Comprehensive Results Summary

Question	Conclusions	Future Directions
Does learning enhance part and part-interaction selectivity? (Aim1, Part 1)	**Yes** • Stronger learned effect for part-interaction • Agreement with Baker et. al. 2002	
Does the learning produce or enhance holistic coding? (Aim1, Part 2)	**No** • Learned selectivity is compositional, not holistic	• Probe familiar part interactions further with genetic algorithm
Are there timing differences between part and part-interaction selectivity? (Aim1, Supplemental)	**Yes** • Main effect precedes Interaction effect by 25-40 ms	
Does learning affect categorical representation? (Aim1, Supplemental)	**Inconclusive** • Too few cells • Categorical representation across trained stimuli trends higher	• More data needed
Do different behavioral contexts affect selectivity? (Aim1, Supplemental)	**Inconclusive** • Too few cells • Task Protocol: 2nd stim selectivity > 1st stim selectivity	• More data needed
Does location affect selectivity or descriptive metrics? (Aim1, Supplemental)	**No** • Could not find any significant trends	
Does the Medial Axis Model fit the data well? (Aim2)	**Modest Performance** • Based faultily on assumption of holistic coding	• Reimagine/Redo • Both the GA and MAM should directly manipulate/investigate multi-part combinatorial neural codes
Does the Medial Axis Model extend Aim1 results? (Aim2)	**Yes: Spatial Contribution** • Unlearned Cells biased toward tops • Learned Cells equalize top versus bottom representation	Extend result to cover GA data

4.1.1 Experiment Design Extensions

Our experiment, stimulus design, and analysis are all inspired by Baker et. al. 2002, but there are notable differences.

- We tasked our monkeys to recognize and differentiate between categories of stimuli instead of single images. In this way, we could ensure that monkeys learned generic shapes, and not rely on ad hoc strategies (involving precise, local details). This is similar to the way we learn alphanumeric characters, which vary in precise shape across fonts and handwriting styles.

- Complementary set, composed of familiar parts but unfamiliar part combinations is unique to our study. This afforded us the ability to directly test changes in part combination selectivity in a way that Baker et. al. could not.

- We applied two-way ANOVA to a larger stimulus set than 2x2 tetrads and directly reported F values. This gave a more direct and more holistic picture of each cell's tuning and selectivity relative to the methodology in Baker et. al.

- We report results from extra analysis / post-hocs, including timing effects, location effects and morph/active context effects.

4.1.2 Aim1, Part1: Confirmation of Learned Increase in Part Selectivity

We observe a learning effect on part selectivity and part interaction selectivity similar to what is reported in Baker et al. Despite the agreement, there is a difference in how the two studies report this. Baker et. al. essentially report counts: that more sessions exhibited significant trained selectivity than non-trained selectivity (modestly so for main effects, drastically so for interaction effects.. We directly report selectivity via F values and

we do so over the entire set of Canonical or Inverted stimuli instead of limiting the analysis to tetrads/sessions. We found a learned effect for bottom and interaction selectivity while top selectivity failed to show a learned difference. Together we interpret this to reproduce the observations of Baker et. al. that learning produced modest parts learning (bottom but not top) and a more significant part interaction learning.

The non-presence of a significant top main effect is difficult to interpret. There is no a priori reason that tops or bottoms should perform differently. One enticing possibility is elucidated later when considering Spatial Contribution: that cells innately represent more information in tops of stimuli rather than bottoms. Through learning, however, cells had to "pay attention" to tops and bottoms equally. Thus, the learning primarily manifests through increased bottom and interaction selectivity.

To play devil's advocate, another speculative possibility involves the relative complexity of our shapes. Our bottom parts tended to either be more complex or occupy more spatial space than the tops. For example, the "cross" and "h" bottom parts not only occupied the lower end, but also the central spatial region of their respective (Unmorphed) stimuli. This is exaggerated further by the simpler top parts, especially the "single-terminator" top. For example the cross letter is essentially just the bottom part.

Furthermore, it is important to note that all of the 2-way ANOVA is done with spatial tops and bottoms as the factors. That means, that the levels of the inverted "tops" correspond to the same medial axis parts as the canonical "bottoms" and thus, inverted "tops" potentially carry more information than inverted "bottoms". Now, given the disproportionate information representation and the "spatial top/bottom" terminology clarification, it is

now possible to reinterpret the main effect graphs. The learned bottom effect is straightforward: canonical bottoms when compared to inverted bottoms, contain more medial axis complexity which works in conjunction with the observed learned effect. However, the canonical tops contain less medial axis complexity than inverted tops, which may counteract any learning effect, thus producing the non-significant relationship we observe. Again, this is all speculative.

4.1.3 Aim1, Part2: Counterevidence to Learned Increase in Holistic Selectivity

We then considered selectivity in Trained versus Complementary responses. This is the key contribution of our experimental design. Baker et. al. observed sharp, sparse tuning in their cells' response profiles and concluded (plausibly) that whole object selectivity explained such tuning. But due to our ability of introducing familiar parts in novel combinations, we could dissociate between a general increase in part interaction selectivity and specific whole-object selectivity, which is what we observed. We recorded multiple cells with sparse response profiles just as Baker et. al. did. However some of those profiles "selected" objects that were never seen in training. Given that Aim 1-Part 1 demonstrates that a learning effect took place (Canonical versus Inverted responses), we conclude then that learning increases general part interaction selectivity, but not whole object selectivity.

This shows that hierarchical part integration is the key step affected by visual object learning. Furthermore, object representation remains fundamentally parts-based even fol-

lowing extensive learning. Parts-based, compositional object coding is efficient enough to comprehend the virtually infinite space of object structures and qualities with a finite number of signals. It is also essential for cognitive understanding of objects as physical structures with interconnected, interacting parts. Face coding also combines information about parts, but in a more promiscuous and implicit way [67], yielding more holistic percepts, with less cognitive and mnemonic access to low-level structural differences. The results reported here argue that compositional coding is maintained even for extremely familiar, behaviorally relevant shapes, preserving coding capacity and explicit structural information about objects.

4.1.4 Selectivity Timing

In a previous study in the lab, Brincat 2006 [56], experimenters reported differential timing effects between linear and nonlinear representations of stimuli in posterior IT cortex. The differentiation of linear and nonlinear components closely resembles the differentiation of main versus interaction components of the current two-way ANOVA. Thus it is worth comparing the two studies and results.

In Brincat 2006, monkeys were shown 2D shape-contour stimuli which could be manipulated in a combinatorial fashion. Neuronal responses were fitted with a model that characterized each neuron's tuning for contour shape (curvature and orientation) and position (x, y object-relative position and x, y absolute position) through time (relative to stimulus onset). The model combined multiple excitatory and inhibitory Gaussian tuning functions (linear components) as well as higher-order products of same-sign tuning functions (non-linear components). In this way, they could delineate linear and nonlinear (modeled)

contributions to neuronal response across time.

The current study differs in obvious and large ways from Brincat 2006. Differences include the stimuli (medial axis definitions versus surface contour), the methodology, and recording location (central/anterior IT versus posterior IT). So comparisons must be made cautiously. Nevertheless, the linear and nonlinear components can be directly compared with the two-way ANOVA executed in the current project. Main effects in two-way ANOVA are essentially linear components in a compositional model of stimuli response, while interaction terms can be interpreted as nonlinear (in that they represent information unable to be captured by linearly summing main effects). That we see a very similar trend of main/linear effects preceding interaction/nonlinear effects (30-40 ms in our study, 60 ms in Brincat 2006) is striking. The convergent results indicate that information about simpler components appears rapidly, whereas information about part interaction or multi-part configurations evolves gradually. Furthermore, it hints at a dynamic transformation from representations of simpler parts to more complex and sparse representation.

4.1.5 Categorical Representation

Within the Morphed protocol we tested categorical representation resulting from training using nested ANOVA. We could not find a significant effect. Comparing nested ANOVA values over Trained versus nested ANOVA values over Complementary responses, for example, failed to show any difference in the ratio of inter-category explained variance to intra-category explained variance. However, more data may be needed to make a definitive conclusion.

4.1.6 Active Versus Passive and Morphed Versus Unmorphed Comparisons

The Morph and Task Protocols, together, comprise a fairly powerful set of post-hocs. For the current project, these experiments were being tweaked throughout the course of recording, and the full version, in which a repeated set of Morphed and Unmorphed stimuli were presented in both passive and active contexts, was only presented to only a handful of cells near the end. Unfortunately there were too few cells to draw significant conclusions. However, interesting trends are reported.

Over the population of cells, Morphed evoked firing rates were modestly larger than Unmorphed responses. A larger effect, however, seemed to occur in comparing active stimuli over passive stimuli. A subset of the cells exhibited striking increases in firing rate in active contexts over passive. It is unclear whether this effect is a neurological effect, or simply due to uncontrolled experimental factors such as cell drift.

Beyond firing rate, there was no significant effect of Morphed or Active effect on one-way ANOVA or sparseness of responses with the exception of Morphed versus Unmorphed active responses: Within the Task protocol, there was greater selectivity (one-way ANOVA) exhibited by the second stimulus (comparison stimulus / unmorphed) than exhibited by the first stimulus (sample stimulus / morphed). This is an interesting effect, and may add to evidence for attentional modulation of ventral visual stream [68–70]. IT activity seems to be "more selective" during the period where active discrimination is taking place. It is worth noting that all trials recorded are from passed trials in which the monkey made the right choice.

4.1.7 Location Effects

We found no notable correlations between location of cell recorded and signal of any kind. We tested two-way ANOVA metrics, one-way Metrics and timing parameters.

4.1.8 Aim2: Modeling Efforts

We utilized a genetic algorithm to explore the shape space of cells The recorded data was used to constrain the Medial Axis Model in order to explain how learned selectivity arises. The Medial Axis Model was the final effort in multiple attempts at modeling the data. Previous attempts included surface contours instead of medial axis components. The final version used templates composed of multiple medial axis elements drawn from top performing stimuli to make a similarity comparison between any stimuli and the general shape characteristics of each cells' best stimuli.

This effort ultimately fell short of potential. Overall model performance, characterized by correlation of observed-versus-predicted firing rates, were modest at best. The technical hurdles centered around the vast combinatorial size of template numbers and template-to-stimuli comparisons proved difficult. Efforts to improve the model were severely limited by time considerations.

But beyond technical limitations, the inclusion of only one template in the Medial Axis Model is fundamentally flawed and any attempts in the future to adapt the model must take this into account (see 4.2 Future Directions). The rationale of modeling with a single template came from the assumption that whole-object selectivity, asserted by Baker et. al., was true. We were not initially attempting to challenge this assertion and the construction

of our model reflects this fact. A single template endeavors to capture a singular shape that the cells have "learned", but in light of our results in Aim 1, which show that selectivity for multiple parts persists throughout learning, models must include multiple templates.

4.1.9 Aim2: Searching for Learned Effect

We developed a learning threshold in order to categorize all recorded cells (even non-Full Protocol cells) as Learned or Unlearned cells. We attempted to differentiate Learned versus Unlearned cells by multiple descriptive metrics. Model Performance and Element Importance failed to show any effect. Spatial Contribution did show an effect (modest, but clear). With, further analysis, we then extend the results of Aim 1 by showing that 1) the Learned cells exhibit an increase of represented top information (not necessarily expected from Aim 1 results) and 2) the primary feature of Learned Cells is the equal top/bottom representation they exhibit over Canonical Stimuli. In other cases (representation of Inverted stimuli by Learned Cells ; representation of all stimuli by Unlearned Cells), cells represented more top information than bottom. This indicates that cells natively "pay attention" more to tops than bottoms, and had to learn to represent information of all parts of trained stimuli.

This analysis covers Medial Axis Model performance over Behavioral Stimuli. An obvious future direction to bolster these results is to also use the performance over genetic algorithm stimuli. That is, top-versus-bottom representation over the genetic algorithm stimuli should be included in future efforts as well. This new analysis should be executed in conjunction with changes to the genetic algorithm and Medial Axis Model.

4.2 Future Directions

4.2.1 New Genetic Algorithm and Template Model

Aim 1 gave us the unexpected and exciting result that part and part interaction selectivity, not whole object selectivity, underlies learned object recognition. Given this conclusion, adjustments should be taken in any future project in this area. Methodologies should adapt to feature parts-based models of the neuron.

First, the genetic algorithm should be modified to include crossover (also called recombination) computations. The current GA only allowed mutations, with the rationale being that we were only interested in whole objects. However, now having established part and part-interaction selectivity as the critical component of neural behavior, the GA should probe this directly. Crossover computations would involve taking two current GA stimuli (two parent stimuli) and "crossing" them to produce a child stimuli. There are many ways this can be executed, but one hypothetical method is outlined as follows. Two parents are randomly selected from the previously run GA (crossover computations can only start from generation 2. Stricter measures could also be put into place by starting crossovers in later generations as top evoked responses become more developed). The selection process would still be weighted by evoked response such that high-response stimuli are more likely to be selected as a parent. However, there would be an added condition that the parent stimuli must have at least two limbs, for reasons that will be obvious shortly. Each parent would randomly produce a "subset skeleton", which are composed of a subset of the skeleton medial axis precursors (see Stimulus Generation Methods, 2.2.2). The skeleton elements are

not final contoured pieces, but the zero-width medial axis precursors) from the respective parent stimuli. Subset skeletons have to satisfy two conditions: 1) all elements must be contiguous, and 2) they must have two to four (inclusive) limbs. The rationale for the lower bound is that recombinations involving single-limb "parts" could theoretically be achieved with a mutation and so nothing would be gained by using a crossover computation. The rationale for the upper bound is that the combined number of limbs of both subset skeletons are subject to the same maximum (six) as normal stimuli. If the number of combined limbs across both subset skeletons exceeds six, then two new subset skeletons will be randomly generated from the same parents. Finally, the two subset skeletons will be combined by randomly selecting one node (a terminator or junction) from each of the subset skeletons and then joining the two stimuli at those nodes to make one final skeleton. Conditions that must be met for this joining step are that the total number of limbs must not exceed four and that the resulting stimuli must be valid (no overlapping limbs). If these conditions are not met, then two new candidate nodes will be selected for joining. Finally, a new contour will be generated around the final skeleton to make a new stimuli.

Another possible change for the genetic algorithm could be the incorporation of parts of the Behavioral stimuli. While it is important for the main component of the genetic algorithm to be independent from the Behavioral stimuli, it is possible that complete independence would task the GA with doing "too much". If the shape space traversed now is expanded to conceptually include multiple parts and their combinations, there may not be enough time to densely sample the space effectively, which is critically important. The Spatial Contribution metrics of Aim 2 strongly indicates robust response to multiple parts

arises as part of learning, and furthermore happens only for familiar parts. So for the genetic algorithm to effectively use crossover mutations to probe part interactions, the parts themselves have to elicit strong reactions. To help alleviate the combinatorial weight, it may be worth considering creating a separate lineage of GA that allows for Behavioral stimuli or parts of the stimuli (i.e. top or bottom parts) to seed new GA stimuli. In this way, the shape space explored would be directly surrounding the Behavioral stimuli (but include a far larger space than that represented by Morphed stimuli) and could be more directly applied to studying part and part interaction tuning.

As the GA methodology changes to probe part and part interactions, the modeling efforts must also change. The Medial Axis Model should incorporate at least two templates and include single and interactive terms. This would not only drastically improve model performance by virtue of more parameters (for example, this would allow for the possibility of having one positive and one negative template which is a fundamental ability that is not possible currently), but would directly follow from the part selectivity results of Aim 1.

The inclusion of a second template in the model was actually an initial goal of the current methodology (when templates was first conceived and implemented), but as previously mentioned, a major technical hurdle was the combinatorial complexity of handling templates. We did not have the computational and time capacity to do everything we wanted, and so the second template was cut from consideration. The assumption then was that whole-object modeling (which can be handled by a single template) could sufficiently describe neural behavior. Now knowing that is not the case, second templates and second order interaction terms are key.

The search for two templates can be done sequentially. After a first template is established, the difference between observed and predicted firing rate would be obtained and the search for the second template would commence fitting to the differential firing rate. The second search would also simultaneously be fitting interaction parameters for the two templates. The time complexity associated with searching with one template is already prohibitive. Adding a second search step is daunting. However it could be alleviated by capping templates at a certain size. Example rules could include templates having no more than 4 limbs, or perhaps having no more than seven total elements. However the time complexity is addressed, the benefits are certainly worth the effort.

4.2.2 New Morph and Task Protocol

The current Morphed stimuli were generated at random, and was not vetted for any sort of criteria. In the future, Morphed stimuli should be parameterized by degree of deviation from unmorphed category shape to ensure a consistency. The exact criteria of deviation will be up to the experimenter. This would be similar in principle to the cat-to-dog morphs that Freedman 2006 [53] generated: a parameterized, repeatable set of stimuli to test with.

The Task Protocol has a balanced design across number of stimuli and trials. This design could be expanded to include active versus passive contexts. Currently, the number of times any given morph appears in a match versus nonmatch trials are random, but it could be balanced if more trials are added (as might be afforded in a chronic recording setting).

4.2.3 New Method: Chronic Recording

One major potential direction for this project would be the incorporation of a chronic array as the data-collection method. Chronic recording gives the experimenter both the ability to hold a cell over multiple weeks and months, as well as the capability to simultaneously record from a population of cells. For this project, the enormous potential benefit would be the ability to observe long-term changes in neural tuning while training is underway. The vast majority of studies concerned with learning (including the current study) only infer that learning has occurred, but do not directly observe learning. The Kobatake 1998 [45] study discussed in Chapter 1 provides one exception, where they take the relatively costly and uncommon step of observing entirely naive control monkeys in parallel with trained monkeys. (for another exception, see Messinger 2001 [71]) By observing neural tuning before, during and after training, the experimenter could now directly ascertain learned signals versus confounds (i.e. differences in naive states across monkeys).

Aside from observing neurons over a long period of time, another benefit to chronic arrays would be population recording. The simultaneous recording of neurons would allow the experimenter to directly study the population coding of visual stimuli. With acute recordings, one can approximately infer population coding (for an example, recall Kobatake 1998 [45], selectivity of trained vs nontrained stimuli was ascertained via population code), but there are problems. Assembling sequentially recorded neurons into a population vector inherently treats each neuron as independent variables. But this may not be the case. Neuronal firing rates may be correlated with each other and representation information may be overlapped. Only through simultaneous population recording can the true population

representation be ascertained. For this project, we consistently deal with questions of discriminability or sparseness of coding across training versus non-training stimuli. We are able to gain a lot of insight (hopefully the reader by this point agrees!) while dealing with these questions on a single cell level, but with chronic arrays, future experimenters could go further.

4.2.3.1 Considerations for a Chronic Array

However, a chronic array does have some drawbacks, some of which are discussed here. In general, the quality of each individual recording channel in a chronic array will be lower than that of each acute recording. With the expectation of capturing multi-unit activity or simply a general decrease in the fidelity of the recorded signal, it may not be possible to study the recorded cells in the same manner as we did with acute recordings. The genetic algorithm and post-hocs run in this study all depended on the quality of signal obtained from single-unit recordings. A move to a chronic array would necessitate a reimagining of the goal and priorities of the recording protocols. Following are considerations for current and new protocols needed for chronic recording.

4.2.3.1.1 Behavioral Protocol

This protocol would be unchanged in terms of execution. The set of Canonical and Inverted stimuli (or new set of trained and untrained stimuli) would be presented in the same fashion (fixation protocol, interleaved trials) as it is currently. What would change is the interpretation / analysis. For acute recordings, the Behavioral Protocol is meant

to ascertain visual responsiveness during the initial search for cells, and only cells that demonstrated suitable shape discriminability were recorded further. For chronic recordings, there will not be a new search for cells day-to-day. Instead, the Behavioral Protocol will serve as a "baseline protocol". The experimenter will compare responses with the previous recording day to ascertain cell drift or loss of signal from the previous day, establishing a "moving baseline" of cell activity that varies by day. This baseline will be critical knowledge when doing analysis of long-term neural activity.

Analysis of the Behavioral Responses should fundamentally involve ascertaining discriminability of Canonical versus Inverted shapes, as the current study does. For Aim 1 of this project we utilized 2-way ANOVA as the main method to ascertain discriminability. For a chronic study, it is possible to repeat the analysis on a channel-by-channel basis, focusing perhaps on a subset of channels with strong signal. To take it further and do a population analysis (taking advantage of simultaneous recording), there exists "multivariate analysis of variance" (MANOVA) which is a generalized form of ANOVA, extending the analysis to two or more dependent variables. In this case signal from each channel is a dependent variable. MANOVA utilizes the covariance matrix across dependent variables which, as mentioned before, is critical for accurately representing a population code. There are a few different versions of MANOVA [72], so further investigation into the appropriate test is warranted.

4.2.3.1.2 Genetic Algorithm

Out of all the protocols, it is the genetic algorithm that stands to gain the most benefit from long-term recordings. In acute recordings, it is a race against time to find a suitable cell and record as many generations of genetic algorithm (along with other protocols) as

possible before the cell drifts. The time constraints limits both the number of generations and number of lineages possible, and for some cells in which the yielded stimuli profile is not very compelling, the experimenter has to accept that the algorithm may simply not have had enough time to find a dynamic region of shape space. The ability to record the same cells (as ascertained by the Behavioral/baseline protocol) over multiple days, however, lifts this major constraint. It will now be possible to "continue" the genetic algorithm from the previous day. Top stimuli can be morphed further while new shapes and lineages can still be added. Furthermore, within a single day, more time can be afforded to the genetic algorithm since without the need to search for cells at the start of each day. All of this extra time allows for a much richer exploration than was possible before.

The objective function (what the algorithm is actually searching) will need to undergo a significant change. Once again, the change from single-cell to population recordings necessitates a change, but in this case, the change is fundamental to the purpose of the protocol. For acute recordings, the genetic algorithm was searching the shape space of a single cell. With a chronic array, though, it would not be feasible to do the same search of the shape space of all the individual channels recorded, both because of the number of cells present and the aforementioned decrease in quality of signal. Instead, the genetic algorithm would be executed over a multivariate signal, in which the "response space" of all (or a subset) of channels are explored at once. But the response space searched will not exactly be the shape space from before, but instead will be something else as framed by the objective function set by the experimenter.

An example objective function could seek to maximize the magnitude of the normalized

response vector of channels, where the normalized response vector would be the response vector multiplied by the inverse of the covariance matrix. This would be a direct multivariate correlate of the shape space search of acute recordings. But this method may not be suitable. Each cell would contribute diffusely to the overall signal and the final product may have a "jack of all trades, master of none" quality, reflecting a little bit of each cell's shape space but none of them directly.

A better objective function may be to look at the sparseness across the population response. Once again, taking advantage of the known sparse coding in IT, this objective function would search for shapes that activate a relatively small number of cells as ascertained by a normalized response vector. An algorithm seeking to maximize sparsity would inherently focus a spotlight on a few cells and thus maximize their responses by honing in on their shape space. Each lineage introduced to the genetic algorithm could potentially hone in on a different set of cells. Different methods of normalization and variants of sparsity exist. One could normalize based on covariance matrix, or baseline firing rate as ascertained by the Behavioral protocol. The experimenter may also wish to attempt to maximize both the sparsity and the maximum response rate. There are a lot of meta parameters to optimize for this new methodology.

4.2.3.1.3 Morphed and Task Protocol

The Morph and Task protocol may have a larger significance in a chronic array setting. For the current study, the collective group of Behavioral, Morph, and Task protocols show all relevant stimuli in exhaustive and balanced manner across canonical versus morphed, and passive versus active settings. Of that group, Morph and Task protocols were post-

hocs that yielded extra information in support of the Behavioral Protocol. For chronic setting, however, one major goal is to record during the training process. As such the Task protocol may become a crucial component. The essential feature of the Morph and Task protocols is the utilization of a set of morphed stimuli that are held constant across all instantiations of the protocols. This should still remain the case. It is important for a consistent set of stimuli to be shown across days for the same "baseline" reasons addressed in the Behavioral Protocol section. However the Task protocol cannot be the only protocol used for training. As explained in the Aim 1 methods, we trained monkeys to recognize categories in order to ensure that it could execute the task with true knowledge of shape characteristics as opposed to rote memorization of arbitrary images. As such, utilizing a repeated set of morphs in the Task Protocol would violate this philosophy. Instead, the Task Protocol should be folded into the existing Training Protocol. Each session should essentially mix "random trials" with "set trials". The random trials are the same as trials from the existing Training protocol: categories picked, match versus nonmatch, and the exact morphed instantiation (if applicable) are all random for any given trial. The "set trials" are from the existing Task Protocol. They are exhaustive and balanced over the number of stimuli and trials. "Random" and "set" trials would be randomly interleaved with each other, with considerably more random trials over set trials. This is to ensure that set trials are infrequent enough that the monkey does not remember the repeated morphs.

Once again, analysis of the data from these protocols would have to be multivariate versions of the current analysis. Both the generalized MANOVA (addressed with Behavioral protocol) and sparsity across population (addressed in the GA protocol) could be useful here.

4.2.3.2 Possible Combination with Acute Recordings

Our task does not favor either side anywhere in the experimental design or monkey saccade direction. Each hemisphere of cortex is treated independently and interchangeably. Thus it is conceptually possible to hold joint chronic and acute recording sessions with the chronic array implanted on one hemisphere, and a separate chamber installed on the other hemisphere. The strategy here, may be to have the chronic array implanted and recording throughout the course of training (multiple months). At this point, the collection of protocols described above would be executed each day. Behavioral, Morphed, and the new Training Protocol described above would collect responses to a repeated set of relevant stimuli, serving as both a baseline test to ascertain cell drift from the previous day, and an updated snapshot of the current "learned state" of IT cortex as training continues. The genetic algorithm will be ongoing throughout the entire training period, with new lineages constantly being added, and old lineages stopped when appropriate.

At the end of training, an acute recording chamber can be installed on the other hemisphere and the current study could be repeated. If possible, joint chronic and acute recordings could happen, though the technical difficulties may be prohibitive (it sounds like a professor's dream and a student/lab technician's nightmare).

4.3 Conclusion

This project showcases a strong continuation of a foundational question of coding in IT cortex. Do IT neurons represent parts or whole objects? We found strong evidence

contradicting the seminal study in the field. Whereas Baker et al. concluded that neurons learn whole, familiar objects, we extended their experimental design and demonstrated that actually, training produces increased interactions between familiar parts without increased selectivity of whole objects. Through a series of post-hocs and modeling efforts, we were able to gain further insights. We found that main part selectivity precedes part interaction selectivity in a strikingly similar way to previously reported results concerning posterior IT activity. We also introduced a Medial Axis Model, through which we gained further insight into the nature of the learning: that naive cells tend to represent information at the spatial tops of stimuli, while learned cells exhibit equal spatial representations, but only across familiar stimuli.

As strong as the findings are, much more work needs to be done. We have found tantalizing clues regarding experience-based neural coding in IT cortex. But to continue further, adjustments must be made to study these clues further and more directly. Efforts to explicitly model learned compositional coding in IT could start with changes to the genetic algorithm to go along with changes to modeling efforts. Results regarding timing differences between main and interaction effects, categorical coding, and attentional effects are all either preliminary or inconclusive with the number of cells recorded. They should be revisited in the future. Finally future work should investigate whether the assertion that naive cells represent more top information is indeed true, as confirmation of this would strengthen the tentative Aim 2 results of this work.

Together, these protocols and results represent a powerful starting point to build off of. I am proud of them and I am excited to see what future versions of them will look like.

Bibliography

[1] N. Savage, "Marriage of mind and machine," *Nature*, vol. 571, no. 7766, pp. S15–17, 2019.

[2] M. A. Goodale and D. Milner, "Separate Visual Pathways for Perception and Action," *Trends in Neurosciences*, vol. 15, no. 1, pp. 20–25, 1992.

[3] M. Mishkin, L. G. Ungerleider, and K. A. Macko, "Object vision and spatial vision: two cortical pathways," *Trends in Neurosciences*, vol. 6, no. C, pp. 414–417, 1983.

[4] K. Grill-Spector, Z. Kourtzi, and N. Kanwisher, "The lateral occipital complex and its role in object recognition," *Vision Research*, vol. 41, no. 10-11, pp. 1409–1422, 2001.

[5] Z. Kourtzi and J. J. Dicarlo, "Learning and neural plasticity in visual object recognition," *Current opinion in neurobiology*, vol. 16, no. 2, pp. 152–158, 2006.

[6] K. L. Hoffman and N. K. Logothetis, "Cortical mechanisms of sensory learning and object recognition," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 364, no. October 2008, pp. 321–329, 2009.

[7] J. L. Gallant, J. Braun, and D. C. Van Essen, "Selectivity for Polar , Hyperbolic , and Cartesian Gratings in Macaque Visual Cortex," *Science*, vol. 259, no. 5091, pp. 100–103, 1993.

[8] B. Y. D. H. Hubel and A. D. T. N. Wiesel, "Receptive Fields, Binocular Interaction And Functional Architecture In The Cat's Visual Cortex," *Journal of Physiology*, vol. 160, pp. 106–154, 1962.

[9] M. Ito and H. Komatsu, "Representation of Angles Embedded within Contour Stimuli in Area V2 of Macaque Monkeys," *Journal of neuroscience*, vol. 24, no. 13, pp. 3313–3324, 2004.

[10] A. Pasupathy and C. E. Connor, "Responses to Contour Features in Macaque Area V4," *Journal of Neurophysiology*, vol. 82, pp. 2490–2502, 1999.

[11] ——, "Population coding of shape in area V4," *Nature Neuroscience*, vol. 5, no. 12, pp. 1332–1338, 2002.

[12] C. Gross, "Visual Cortex Properties of Neurons in Inferotemporal of the Macaque," *Journal of neurophysiology*, vol. 35, no. 1, pp. 96–111, 1972.

[13] R. Desimone and C. G. Gross, "Visual areas in the temporal cortex of the macaque," *Brain Research*, vol. 178, pp. 363–380, 1979.

[14] R. Desimone, T. D. Albright, C. G. Gross, and C. Bruce, "STIMULUS-SELECTIVE NEURONS IN THE MACAQUE1 OF INFERIOR TEMPORAL," *Journal of Neuroscience*, vol. 4, pp. 2051–2062, 1984.

[15] D. Y. Tsao, "A Cortical Region Consisting Entirely of Face-Selective Cells," *Science*, vol. 15, no. May, p. 58, 2006.

[16] J. Richmond and H. Wurtz, "Visual Responses of Inferior Temporal Neurons in Awake Rhesus Monkey," *Journal of Neurophysiology*, vol. 50, no. 6, pp. 1415–1432, 1983.

[17] D. A. Pollen, M. Nagler, J. Daugman, R. Kronauer, and P. Cavanagh, "USE OF GABOR ELEMENTARY FUNCTIONS TO PROBE RECEPTIVE FIELD SUBSTRUCTURE OF POSTERIOR," *Vision Research*, vol. 21, pp. 233–241, 1984.

[18] R. Vogels and G. A. Orban, "Activity of Inferior Temporal Neurons During Orientation Discrimination With Successively Presented Gratings," *Journal of Neurophysiology*, vol. 71, no. 4, pp. 1428–1451, 1994.

[19] D. I. Perrett, E. T. Rolls, and W. Caan, "Visual neurones responsive to faces in the monkey temporal cortex," *Experimental brain research.*, vol. 47, pp. 329–342, 1982.

[20] E. T. Rolls and G. C. Baylis, "Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey," *Experimental Brain Research.*, vol. 65, pp. 38–48, 1986.

[21] S. Yamane, S. Kaiji, and K. Kawano, "What facial features activate face neurons in the inferotemporal cortex of the monkey?" *Experimental Brain Research.*, vol. 73, pp. 209–214, 1988.

[22] M. E. Hasselmo, E. T. Rolls, G. C. Baylis, and V. Nalwa, "Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey," *Experimental Brain Research.*, vol. 75, pp. 417–429, 1989.

[23] M. P. Young and S. Yamane, "Sparse Population Coding of Faces in the Inferotemporal Cortex," *Science*, vol. 256, no. 5061, pp. 1327–1331, 1992.

[24] T. Sato, T. Kawamura, and E. Iwai, "Responsiveness of inferotemporal single units to visual pattern stimuli in monkeys performing discrimination," *Experimental Brain Research*, vol. 38, no. 3, pp. 313–319, 1980.

[25] G. Sary, R. Vogels, and G. A. Orban, "Cue-Invariant Shape Selectivity of Macaque Inferior Temporal Neurons," *Science*, vol. 260, no. 5110, pp. 995–997, 1993.

[26] C. I. Baker, M. Behrmann, and C. R. Olson, "Impact of learning on representation of parts and wholes in monkey inferotemporal cortex," *Nature Neuroscience*, vol. 5, no. 11, pp. 1210–1216, 2002.

[27] D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller, "A Comparison of Primate Prefrontal and Inferior Temporal Cortices during Visual Categorization," *Journal of Neuroscience*, vol. 23, no. 12, pp. 5235–5246, 2003.

[28] I. Fujita, K. Tanaka, M. Ito, and K. Cheng, "Columns for visual features of objects in monkey inferotemporal cortex," *Letters to Nature*, vol. 360, pp. 343–346, 1992.

[29] N. K. Logothetis, J. Pauls, and T. Poggio, "Shape representation in the inferior temporal cortex of monkeys," *Current Biology*, vol. 5, no. 5, pp. 552–563, 1995.

[30] K. Tsunoda, Y. Yamane, M. Nishizaki, and M. Tanifuji, "Complex objects are represented in macaque inferotemporal cortex by," *Nature neuroscience*, vol. 4, no. 8, pp. 832–838, 2001.

[31] M. Mishkin and K. H. Pribram, "Visual Discrimination Performance Following Partial Ablations," *Journal of Comparative and Physiological Psychology*, vol. 47, no. 1, pp. 14–20, 1954.

[32] M. Mishkin, "Visual discrimination performance following partial ablations of the temporal lobe: II. Ventral surface vs. hippocampus," *Journal of Comparative and Physiological Psychology*, vol. 47, no. 3, pp. 187–193, 1954.

[33] J. Horel and L. J. Misantone, "Visual Discrimination Impaired by Cutting Temporal Lobe Connections," *Science*, vol. 193, pp. 336–338, 1976.

[34] J. A. Horel, D. E. Pytko-Joiner, M. L. Voytko, and K. Salsbury, "The performance of visual tasks while segments of the inferotemporal cortex are suppressed by cold," *Behavioural Brain Research*, vol. 23, no. 1, pp. 29–42, 1987.

[35] J. A. Horel, "Retrieval of a face discrimination during suppression of monkey temporal cortex with cold," *Neuropsychologia*, vol. 31, no. 10, pp. 1067–1077, 1993.

[36] S. R. Afraz, R. Kiani, and H. Esteky, "Microstimulation of inferotemporal cortex influences face categorization," *Nature*, vol. 442, no. 7103, pp. 692–695, 2006.

[37] R. M. Seyfarth, D. L. Cheney, and P. Marler, "Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication," *Science*, vol. 210, no. 4471, pp. 801–803, 1980.

[38] H. R. Rodman, "Development of inferior temporal cortex in the monkey," *Cerebral Cortex*, vol. 4, no. 5, pp. 484–498, 1994.

[39] C. Blakemore and R. C. Van Sluyters, "Innate and environmental factors in the development of the kitten's visual cortex." *The Journal of Physiology*, vol. 248, no. 3, pp. 663–716, 1975.

[40] J. V. Bouvrie and P. Sinha, "Visual object concept discovery: Observations in congenitally blind children, and a computational approach," *Neurocomputing*, vol. 70, no. 13, pp. 2218–2233, 2006.

[41] C. D. Gilbert, M. Sigman, and R. E. Crist, "The neural basis of perceptual learning," *Neuron*, vol. 31, no. 5, pp. 681–697, 2001.

[42] P. G. Schyns, R. L. Goldstone, and J. P. Thibaut, "The development of features in object concepts," *Behavioral and Brain Sciences*, vol. 21, no. 1, pp. 1–54, 1998.

[43] D. Herrnstein, R.J., Loveland, "Complex Visual Concept in the Pigeon," *Science*, vol. 146, pp. 549–551, 1964.

[44] M. A. Qadri and R. G. Cook, "Pigeons and humans use action and pose information to categorize complex human behaviors," *Vision Research*, vol. 131, pp. 16–25, 2017.

[45] E. Kobatake, G. Wang, and K. Tanaka, "Effects of Shape-Discrimination Training on the Selectivity of Inferotemporal Cells in Adult Monkeys," *Journal of Neurophysiology*, vol. 80, no. 1, pp. 324–330, 1998.

[46] M. Fahle, "Perceptual learning: A case for early selection," *Journal of Vision*, vol. 4, no. 10, pp. 879–890, 2004.

[47] I. Fine and R. A. Jacobs, "Comparing perceptual learning tasks: A review," *Journal of Vision*, vol. 2, no. 2, pp. 190–203, 2002.

[48] G. H. Recanzone, M. M. Merzenich, and W. M. Jenkins, "Frequency discrimination training engaging a restricted skin surface results in an emergence of a cutaneous response zone in cortical area 3a," *Journal of Neurophysiology*, vol. 67, no. 5, pp. 1057–1070, 1992.

[49] G. H. Recanzone, C. E. Schreiner, and M. M. Merzenich, "Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys," *Journal of Neuroscience*, vol. 13, no. 1, pp. 87–103, 1993.

[50] A. Schoups, R. Vogels, N. Qian, and G. Orban, "Practising orientation identification improves orientation coding in V1 neurons," *Nature*, vol. 412, no. 6846, pp. 549–553, 2001.

[51] T. Yang and J. H. Maunsell, "The Effect of Perceptual Learning on Neuronal Responses in Monkey Visual Area V4," *Journal of Neuroscience*, vol. 24, no. 7, pp. 1617–1626, 2004.

[52] E. Zohary, S. Celebrini, K. H. Britten, and W. T. Newsome, "Neuronal Plasticity that Underlies Improvement in Perceptual Performance Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/2883477," vol. 263, no. 5151, pp. 1289–1292, 1994.

[53] D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller, "Experience-Dependent Sharpening of Visual Shape Selectivity in Inferior Temporal Cortex," no. November, pp. 1631–1644, 2006.

[54] K. Saleem and N. K. Logothetis, *A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates*, 2nd ed. Academic Press, 2012.

[55] E. T. Carlson, R. J. Rasquinha, K. Zhang, and E. Charles, "NIH Public Access," *Current Biology*, vol. 21, no. 4, pp. 288–293, 2011.

[56] S. L. Brincat and C. E. Connor, "Dynamic Shape Synthesis in Posterior Inferotemporal Cortex," *Neuron*, vol. 49, pp. 17–24, 2006.

[57] K. G. Thompson, D. P. Hanes, N. P. Bichot, and J. D. Schall, "Perceptual and Motor Processing Stages Identified in the Activity of Macaque Frontal Eye Field Neurons During Visual Search," *Journal of neurophysiology*, vol. 76, no. 6, pp. 4040–4055, 1996.

[58] S. L. Brincat and C. E. Connor, "Underlying principles of visual shape selectivity in posterior inferotemporal cortex." *Nature neuroscience*, vol. 7, no. 8, pp. 880–6, aug 2004.

[59] Y. Yamane, E. T. Carlson, K. C. Bowman, Z. Wang, and C. E. Connor, "A neural code for three-dimensional object shape in macaque inferotemporal Cortex," *Nature Neuroscience*, vol. 11, no. 11, pp. 1352–1360, 2008.

[60] C. C. Hung, E. T. Carlson, and C. E. Connor, "Medial Axis Shape Coding in Macaque Inferotemporal Cortex," *Neuron*, vol. 74, no. 6, pp. 1099–1113, 2012.

[61] H. Blum, "Biological shape and visual science (part I)," *Journal of Theoretical Biology*, vol. 38, no. 2, pp. 205–287, 1973.

[62] D. Marr and H. Nishihara, "Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes Source," *Proceedings of the Royal Society of London*, vol. 200, no. 1140, pp. 269–294, 1978.

[63] B. B. Kimia, "On the role of medial geometry in human vision," *Journal of Physiology Paris*, vol. 97, pp. 155–190, 2003.

[64] C. Arcelli, L. P. Cordella, and S. Levialdi, "From Local Maxima to Connected Skeletons," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 3, no. 2, pp. 134–143, 1981.

[65] S. M. Pizer, W. R. Oliver, and S. H. Bloomberg, "Hierarchical Shape Description Via the Multiresolution Symmetric Axis Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 505–511, 1987.

[66] M. F. Demirci, A. Shokoufandeh, and S. J. Dickinson, "Skeletal shape abstraction from examples," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 944–952, 2009.

[67] L. Chang and D. Y. Tsao, "The Code for Facial Identity in the Primate Brain Article The Code for Facial Identity in the Primate Brain," *Cell*, vol. 169, no. 6, pp. 1013–1020.e14, 2017.

[68] D. Felleman and D. Van Essen, "Distributed Hierachical Processing in the Primate Cerebral Cortex," *Cerebral Cortex*, vol. 1, pp. 1–47, 1991.

[69] J. Bullier, J.-m. Hup, A. C. James, and P. Girard, "The role of feedback connections in shaping the responses of visual cortical neurons," *Progress in Brain Research*, vol. 134, pp. 1–12, 2001.

[70] H. Stemmann and W. A. Freiwald, "Evidence for an attentional priority map in inferotemporal cortex," *Proceedings of the National Academy of Sciences*, vol. 116, no. 47, pp. 23 797–23 805, 2019.

[71] A. Messinger, L. R. Squire, S. M. Zola, and T. D. Albright, "Neuronal representations of stimulus associations develop in the temporal lobe during learning," *Proceedings of the National Academy of Sciences*, vol. 98, no. 21, pp. 12 239–12 244, 2001.

[72] R. Warne, "Practical Assessment , Research , and Evaluation A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists," *Practical Assessment, Research, and Evaluation*, vol. 19, pp. 1–10, 2014.

Vita

Andrew Cheng attended the University of Michigan where he first found a love for research. Taking a lead role in an undergraduate synthetic biology team, he led an effort to reprogram Ecoli cells to be digital binary counters. Along the way he discovered the wonderful work of mathematical modeling. In 2009, upon graduation with a degree in Biomedical Engineering, he enrolled at Johns Hopkins University to pursuing a PhD, also in Biomedical Engineering. Here, Andrew transitioned to neuroscience research, specializing in first somatosensory, then visual systems neuroscience. Along the way, Andrew gained an exciting, multidisciplinary experience involving prosthetic/haptic feedback arms, development of low cost medical devices, and designing multi-electrode arrays. One of the highlights of his graduate

(and undergraduate) experience has been the role of teaching. Andrew thoroughly enjoyed his time as a Teaching Assistant. Holding office hours, working with students, and grading (yes...even grading) were tasks that most grad students despised, but Andrew loved every minute of it! Currently, upon graduation, Andrew plans to transition to becoming a STEM high school teacher in an underserved community. Outside of academic endeavors, Andrew is a high energy person with a never-sit-down, up-for-anything mindset. Hobbies include running, climbing, swing dancing, yoga, basketball, and weight lifting. His spontaneity and unpredictability has been known to drive his wife crazy from time to time, but they love each other very much! (He thinks)